Human Centered Multimedia
Institute of Computer Science

Universität Augsburg University

Analyse experimenteller Daten

-

Häufigkeiten und Korrelationen

Human Centered Multimedia
Institute of Computer Science
Augsburg University
Universitätsstr. 6a
86159 Augsburg, Germany

Arten von Tests

Häufigkeiten (nominalskalierte Daten)
oder
Bewertungen, Messungen (ordinal- oder intervallskalierte Daten)?

Häufigkeiten

Bewertungen, Messungen

Experimentelles oder korrelatives Design?

Experiment
(Unterschiede zwischen Gruppen oder Konditionen)
(siehe vorherige Foliensätze)

Korrelatives Design
(Beziehungen zwischen zwei unabhängigen Variablen)

Häufigkeiten - chi^2 Test (eindimensional)

- Test von Hypothesen über die Verteilung einer **nominalskalierten Variablen** in einer Population
- Verteilung absoluter Häufigkeiten: Kategorisierung der Versuchspersonen hinsichtlich des nominalskalierten Merkmals M mit k Stufen

- Beispiel: Verwendung einer Lernsoftware nach Geschlecht

Männer	Frauen	Σ
40	120	160

Häufigkeiten -
chi² Test (eindimensional)

- Ermittelt, welche Häufigkeitsverteilung in der Population gilt, aus der die Stichprobe stammt.
- Gegenüberstellung der theoretischen Annahme über die Häufigkeits-verteilung mit der tatsächlich beobachteten Häufigkeitsverteilung

Nullhypothese:

Die theoretisch erwartete Häufigkeitsverteilung entspricht der beobachteten (empirisch ermittelten) Häufigkeitsverteilung.

Alternativhypothese:

Es gibt einen systematischen Unterschiede zwischen der theoretisch angenommenen und der empirisch ermittelten Häufigkeitsverteilung.

Häufigkeiten - chi² Test (eindimensional)

- Ziel: Widerlegen der Nullhypothese, um die Alternativhypothese zu bestätigen.

- Unterschied zu t-Test:
  - Jede begründete Annahme über die Auftrittswahrscheinlichkeiten kann als Nullhypothese fungieren.
    - Verschiedene Nullhypothese möglich
  - Naheliegende Nullhypothese: Gleichverteilung aller Merkmalsstufen (Gleichverteilungshypothese)
  - Besser: Nullhypothese basierend auf theoretischen Überlegungen und vorhandenen Statistiken

Häufigkeiten - chi² Test (eindimensional)

Gleichverteilungsannahme:

- Häufigkeit über alle Stufen des Merkmals hinweg gleich

$$f_{e1} = f_{e2} = \ldots = f_{ek} = \frac{N}{k}$$

$$f_{ei} \quad : \quad erwartete\ Häufigkeit\ in\ Kategorie\ i$$

$$k \quad : \quad Anzahl\ der\ Kategorien\ des\ Merkmals$$

$$N \quad : \quad Stichprobe\,numfang$$

- Beispiel (Verwendung der Lernsoftware):
  Σ = 160; Kategorien: 1. Frauen, 2. Männer

$$f_{e1} = f_{e2} = \frac{160}{2} = 80$$

Häufigkeiten - chi² Test (eindimensional)

chi² Kennwert

- Misst die Abweichung der beobachteten von den erwarteten Häufigkeiten.
- Kennwert für die Entscheidung über statistisch signifikante Abweichungen
- Folgt der charakteristischen chi²-Verteilung (siehe Abbildung), die abhängig von den Freiheitsgraden ist (df = k – 1)

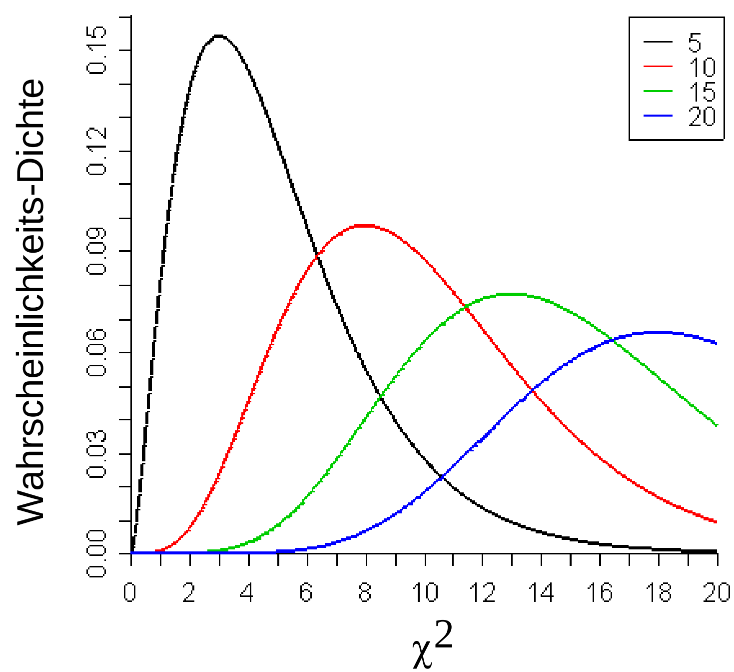
- Formel:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_{bi} - f_{ei}\right)^2}{f_{ei}}$$

k Anzahl der Kategorien des Merkmals

f_{bi} beobachtete Häufigkeit in Kategorie i

f_{ei} erwartete Häufigkeit in Kategorie i



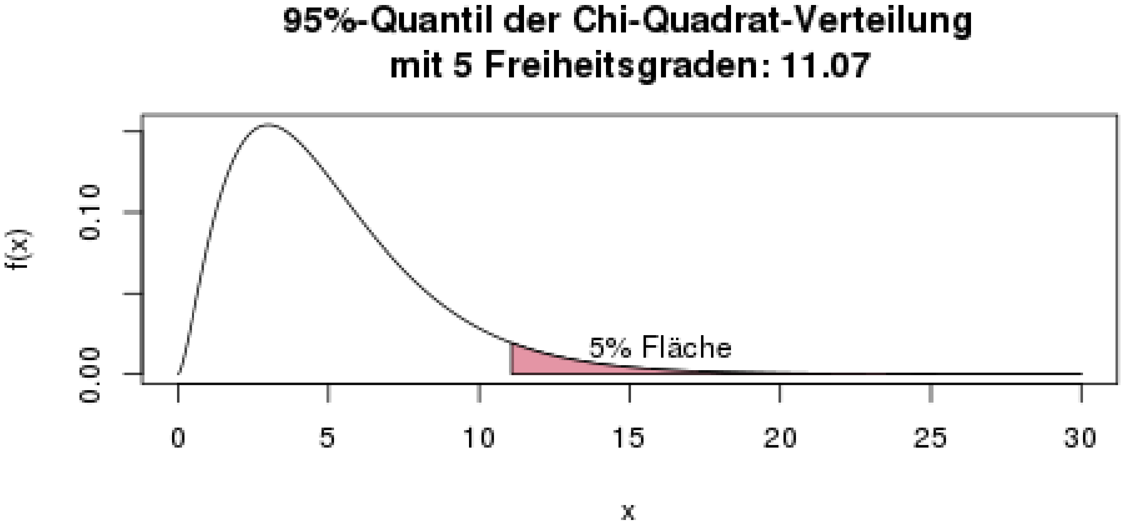
Bestimmung der Freiheitsgrade df

- Anzahl der Häufigkeiten, die unabhängig voneinander variieren können.
- Summe der beobachteten Häufigkeiten für k betrachtete Kategorien muss immer N (= Stichprobengröße) ergeben.
  - k-1 Summanden können variieren => df = k-1

Universität Augsburg University

Signifikanzprüfung

- Jedem chi^2-Wert ist in Abhängigkeit von seinen Freiheitsgraden eine Wahrscheinlichkeit zugeordnet.
- Wahrscheinlichkeit des gefundenen chi^2-Wert oder eines größeren chi^2-Werts unter der Nullhypothese
- Ist diese Wahrscheinlichkeit kleiner als ein vorher festgelegtes Signifikanzniveau α, so wird die Nullhypothese verworfen und die Alternativhypothese angenommen.



Häufigkeiten - chi² Test (eindimensional)

Beispiel:

- Verwendung der Lernsoftware in Abhängigkeit vom Alter

	Kinder	**Jugendliche**	**Erwachsene**
f_{bi}	62	155	83
f_{ei}	90	150	60
$f_{bi}-f_{ei}$	-28	5	23
$(f_{bi}-f_{ei})^2$	784	25	529
$(f_{bi}-f_{ei})^2/f_{ei}$	8,71	0,17	8,82

$$df = 3 - 1 = 2$$

$$\chi^2 = 8{,}71 + 0{,}17 + 8{,}82 = 17{,}7$$

df	**0,95**	**0,99**	**0,999**
1	3,841	6,635	10,828
2	5,991	9,210	**13,816**

- Der errechnete Wert liegt bei df = 2 über dem kritischen Wert in der Tabelle bei einem Signifikanzniveau von α=0,001
- ➢ Es liegt ein höchst signifikanter Unterschied vor.

Häufigkeiten - Gerichtetes Testen

- Der eindimensionale chi^2-Test ist prinzipiell ungerichtet.
- Bei genau zwei Kategorien ist auch ein gerichtetes Testen möglich.
- Nur in diesem Fall ist eindeutig zwischen welchen Stufen das signifikante Ergebnis aufgetreten ist.
- Gerichtete Hypothese: „Kategorie A tritt häufiger auf als Kategorie B"
- Verdopplung des gewünschten Signifikanzniveau α (z.B. 5% -> 10%)
  - ➢ Verringerung des kritischen chi2-Werts
  - ➢ Ergebnis wird eher signifikant als bei zweiseitiger Fragestellung
  - ➢ Teststärke nimmt zu

Beispiel (Soziales Netzwerk für Computer Spiele):

- 824 Mitglieder (441 männlich, 383 weiblich)
- Gerichtete Hypothese:
  In diesem Netzwerk sind eher Männer als Frauen registriert.
- Gleichverteilungshypothese: Geschlechterverhältnisse 1:1 (je 412)

$$\chi^2 = \frac{(441-412)^2}{412} + \frac{(383-412)^2}{412} = 4{,}08$$

df	0,9
1	**2,706**

- Doppeltes α-Niveau von 10%
  - ➢ kritischer chi²-Wert (df=1, α=0,1) beträgt 2,706.
- Der errechnete Wert liegt über dem kritischen chi²-Wert.

➢ In diesem Netzwerk registrieren sich signifikant mehr Männer als Frauen.

Häufigkeiten - chi²-Test (zweidimensional)

- chi^2-Test mit zwei nominalskalierten Merkmalen mit mindestens zwei Stufen
- Zuordnung der Versuchspersonen zu den möglichen Kombinationen der Stufen beider Merkmale
- Kreuztabelle:
  - Zeilen: Merkmal A mit k-Stufen
  - Spalten: Merkmal B mit l Stufen
- Der zweidimensionale chi^2-Test heißt deshalb auch k x l-chi^2-Test.
- Besondere Form des zweidimensionalen chi^2-Tests: Kontingenzanalyse = Test auf Unabhängigkeit der beiden Merkmale

Häufigkeiten - Kontingenzanalyse

- Ermöglicht Aussage, ob zwei betrachtete Merkmale in irgendeiner Form stochastisch zusammenhängen.

- Beispiele für Forschungsfragen:
  - Ist die Präferenz für einen bestimmten Roboter abhängig vom Geschlecht?
  - Bevorzugen Versuchspersonen eher Roboter, die dasselbe Geschlecht wie sie selber haben?

- Wesentlicher Unterschied zum eindimensionalen Test:
  Nicht die Verteilung der Merkmale als solche, sondern die Beziehung der beiden Merkmale zueinander steht im Vordergrund.

Häufigkeiten - Kontingenzanalyse

Nullhypothese:

Die beiden Merkmale sind stochastisch unabhängig voneinander.

Alternativhypothese:

Zwischen den Stufen des einen Merkmals und des Stufen des anderen besteht ein irgendwie gearteter Zusammenhang.

- Anhand der Annahme der Unabhängigkeit beider Merkmale lassen sich die erwarteten Zellhäufigkeiten schätzen.
- Diese werden dann mit den beobachteten Zellhäufigkeiten verglichen.
- Der chi^2-Wert fungiert als Maß für die Abweichung der beobachteten von den erwarteten Werten.

Berechnung der erwarteten Häufigkeiten unter der H0

Erhobene Daten in Form einer Kreuz- oder Kontingenztabelle

	B1	**B2**	**…**	**Bj**	**Σ**
A1	n11	n12	…	n1j	**n1.**
A2	n21	n22	…	n2j	**n2.**
…	…	…	…	…	**…**
Ai	ni1	ni2	…	nij	**ni.**
Σ	**n.1**	**n.2**	**…**	**n.j**	**N**

Häufigkeiten -
chi^2-Test (zweidimensional)

Beispiel:

- Hängt die Einstellung zu Robotern mit dem Alter zusammen?

	Kinder	**Jugendliche**	**Erwachsene**	
Befürwortung	28	16	6	**50**
Ablehnung	32	64	54	**150**
	60	**80**	**60**	**200**

- Insgesamt befürworten 50 Personen den Einsatz von Robotern im Haushalt, 150 lehnen ihn ab.
- Damit liegt die Wahrscheinlichkeit, dass Personen der drei Altersgruppen den Einsatz von Robotern befürworten bzw. ablehnen bei 50/200 = 25% bzw. 150/200 = 75%.

Häufigkeiten - chi²-Test (zweidimensional)

Beispiel:

- Das Verhältnis von Befürwortung und Ablehnung beträgt 25% zu 75%.
- Dieses Verhältnis müsste sich auch in den einzelnen Gruppen wiederfinden.

		Kinder	Jugendliche	Erwachsene	
Befürwortung	**fb**	28	16	6	**50**
	fe	15	20	15	**50**
Ablehnung	**fb**	32	64	54	**150**
	fe	45	60	45	**150**
		60	**80**	**60**	**200**

- Die beobachteten Werte weichen sichtbar von den unter der Nullhypothese der Unabhängigkeit erwarteten Werten ab.
- Die Frage ist jedoch, ob die Diskrepanzen statistisch bedeutsam sind oder auf zufälligen Schwankungen beruhen.

Universität Augsburg University

Der chi² Kennwert

- Dient zur Entscheidung über statistisch signifikante Unterschiede zwischen den erwarteten und den beobachteten Häufigkeiten
- Die Formel ähnelt der für den eindimensionalen Fall:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(f_{bij} - f_{eij}\right)^2}{f_{eij}}$$

k : *Anzahl der Kategorien des Merkmals A*

l : *Anzahl der Kategorien des Merkmals B*

f_{bij} : *beobachtete Häufigkeit der Merkmalskombination i, j*

f_{eij} : *erwartete Häufigkeit der Merkmalskombination i, j*

- Wert für das Beispiel: $\chi^2 = 23{,}29$

Berechnung der Freiheitsgrade:

- Folgende Überlegung:
  - Die Addition der Zeilensummen muss N ergeben.
  - Hier können k-1 Summanden frei variieren.
  - Ebenso müssen sich die Spaltensummen zu N ergeben.
  - Hier können l-1 Summanden frei variieren.
- Da Zeilen- und Spaltensumme auf dieselben Werte zurückgreifen, ergibt sich die Anzahl der Freiheitsgrade beim zweidimensionalen chi^2-Test zu: $df = (k-1) \cdot (l-1)$

- Im Beispiel: $df = (3-1) \cdot (2-1) = 2$

Signifikanzprüfung:

- Die chi^2-Verteilung assoziiert jeden chi^2-Wert bei gegebener Anzahl der Freiheitsgrade mit einer bestimmten Wahrscheinlichkeit.
- Wahrscheinlichkeit der beobachteten Häufigkeitsverteilung unter Annahme der Nullhypothese (Unabhängigkeit)
- Je größer der chi^2-Wert, um unwahrscheinlicher ist es, dass in der Population, aus der die Stichprobe stammt, die Nullhypothese gilt.
- Ist die Wahrscheinlichkeit kleiner als ein zuvor festgelegtes Signifikanzniveau, so wird die Nullhypothese verworfen.

- Beispiel: df = 2, kritischer Wert = 13.816 (α = 0,001). $\chi^2 = 23{,}29$
  - Die Einstellung zum Einsatz von Robotern hängt mit dem Alter zusammen.

df	**0,95**	**0,99**	**0,999**
1	3,841	6,635	10,828
2	5,991	9,210	**13,816**

Häufigkeiten - Zusammenfassung

- chi^2 gibt Information darüber, ob sich die erwarteten Häufigkeiten von den beobachteten Häufigkeiten unterscheiden.
  - Keine Information über mögliche Gründe
- Zweidimensionaler chi^2 gibt Information darüber, ob die untersuchten Variablen stochastisch korrelieren.
  - Keine Information über mögliche kausale Beziehungen
- Üblicherweise, aber nicht immer, kann man anhand der Diskrepanzen zwischen den beobachteten und den erwarteten Häufigkeiten erkennen, warum der Chi^2-Test zu einem signifikanten Ergebnis führte.
- Oft sind die Ergebnisse eines Chi^2-Tests nicht einfach zu interpretieren.
- Jede Versuchsperson muss genau einer Kategorie zuzuordnen sein. Andernfalls, sind die Ergebnisse komplett wertlos.

Beispiel - Interpretation:

- Die beobachteten Werte weichen signifikant von den unter der Nullhypothese der Unabhängigkeit erwarteten Werten ab.
- Es macht keinen Sinn, dass die Meinung über Roboter das Alter der Teilnehmer beeinflusst.
  - Die Zuneigung gegenüber Robotern hängt vom Alter ab.
- Kinder sind Robotern aufgeschlossener gegenüber Robotern, als Jugendliche und Erwachsene.

		Kinder	Jugendliche	Erwachsene	
Befürwortung	**fb**	28	16	6	**50**
	fe	15	20	15	**50**
Ablehnung	**fb**	32	64	54	**150**
	fe	45	60	45	**150**
		60	**80**	**60**	**200**

Arten von Tests

Ziel: Häufigkeiten oder Bewertungen?

Häufigkeiten

Bewertungen

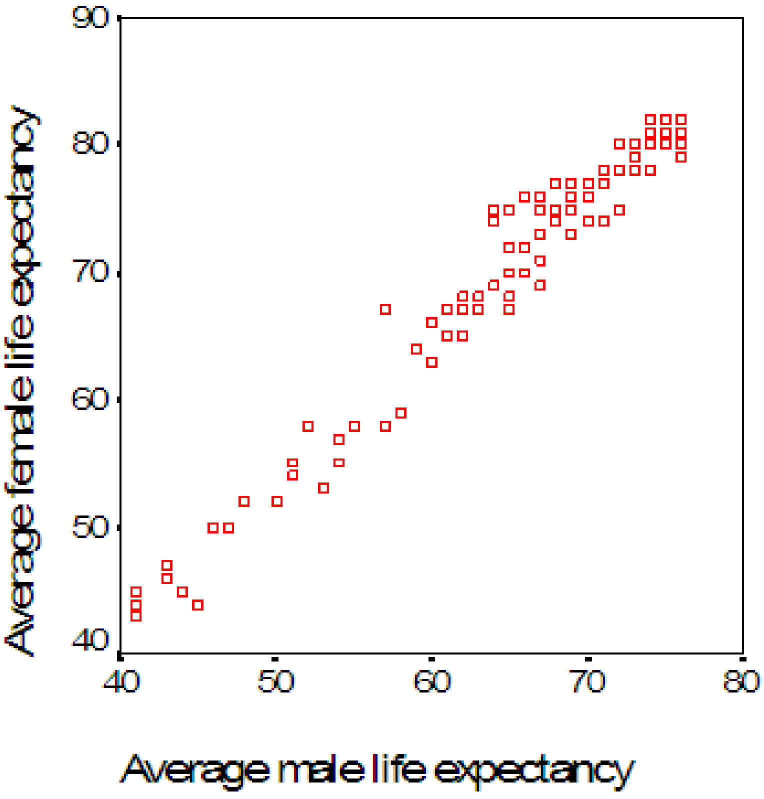
Experimentelles oder korrelatives Design?

Experiment
(Unterschiede zwischen Gruppen oder Konditionen)
(siehe vorherige Foliensätze)

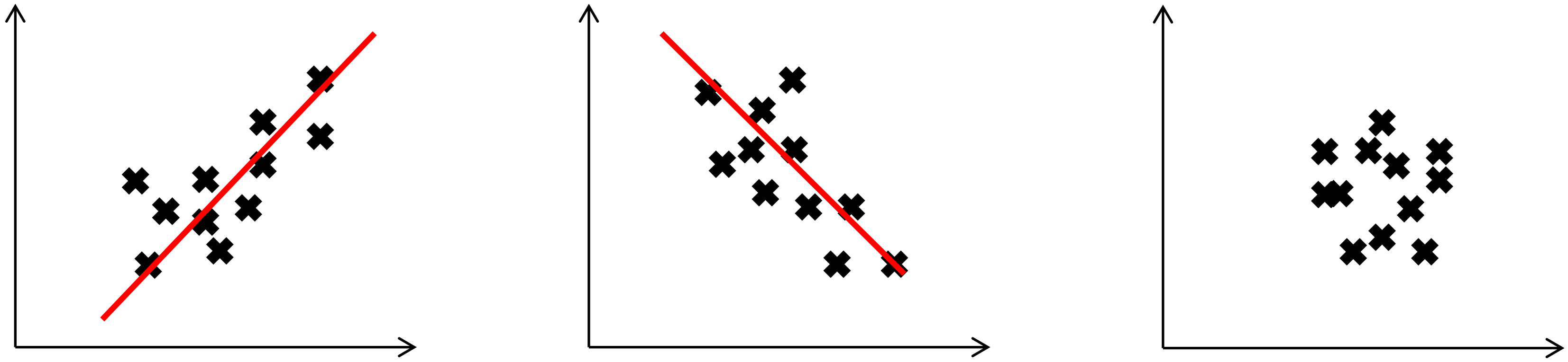
Korrelatives Design
(Beziehungen zwischen zwei unabhängigen Variablen)

Korrelationen

- Gibt Auskunft über die Stärke des Zusammenhangs zwischen zwei Variablen.
- Stärke des Zusammenhangs wird durch den Korrelationskoeffizienten r wiedergegeben.
- Der Korrelationskoeffizient kann zwischen –1 und +1 liegen.
- Von einer Korrelation kann NICHT auf einen kausalen Zusammenhang geschlossen werden.



- Der Grad des Zusammenhangs zwischen zwei **intervallskalierten Variablen** lässt sich mathematisch durch die **Kovarianz** und die auf ihr aufbauende **Produkt-Moment-Korrelation** beschreiben.
- Hängen zwei Variablen zusammen, so gibt die Ausprägung, die eine Versuchsperson auf der einen Variablen aufweist, zu gewissen Teilen auch Auskunft darüber, welche Ausprägung diese Person auf der anderen Variablen erreicht.



Positiver Zusammenhang | Negativer Zusammenhang | Kein Zusammenhang

Korrelationen – Kovarianz

- Standardisiertes Maß, das den Zusammenhang zweier Variablen erfasst.
- Im Gegensatz zur Varianz macht die Kovarianz Aussagen über die gemeinsame Variation zweier Merkmale:

$$cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1}$$

- Positive Kovarianz: Beide Variablen weichen weitgehend gemeinsam in die gleiche Richtung von ihrem Mittelwert ab.
- Negative Kovarianz: Es treten viele entgegengesetzte Abweichungen vom jeweiligen Mittelpunkt auf.

Korrelationen –
Produkt-Moment-Korrelation

- Gebräuchlichstes Maß für die Korrelation zweier Variablen: Produkt-Moment-Korrelation nach Pearson
- Ausgedrückt im Korrelationskoeffizienten r:
  - Standardisierung der Kovarianz
  - Relativierung der empirisch ermittelte Kovarianz an der maximalen Kovarianz
  - Wertebereich: -1 bis 1

$$r = \frac{\mathrm{cov}_{emp}}{\mathrm{cov}_{\max}} = \frac{\mathrm{cov}(x, y)}{\hat{\sigma}_x . \hat{\sigma}_y}$$

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)\hat{\sigma}_x \hat{\sigma}_y} = \frac{1}{n-1} \sum_{i=1}^{n} \frac{x_i - \bar{x}}{\hat{\sigma}_x} \frac{y_i - \bar{y}}{\hat{\sigma}_y}$$

Entspricht den Formeln für z-Standardisierung angewandt auf beide Variablen

Korrelationen – Kausalität

- **Achtung:** Korrelationen dürfen nicht vorschnell als Kausalbeziehung betrachtet werden.
- Mögliche Abhängigkeit von einer dritten gemeinsamen Variable $\Rightarrow$ Scheinkorrelation
  - Nach dem ersten Weltkrieg nahmen sowohl die Anzahl der Geburten als auch die Anzahl der Störche ab.
  - ABER: Beide Rückgänge können plausibler als Folgen der zunehmenden Industrialisierung betrachtet werden.
- Richtung der Beeinflussung ist unklar
  - Es wurde empirisch ein Zusammenhang zwischen dem Konsum gewaltverherrlichender Spiele und der Aggressionsbereitschaft von Personen ermittelt.
  - ABER: Wird die Aggressionsbereitschaft durch gewaltverherrlichende Spiele verursacht, oder spielen aggressive Leute eher solche Spiele?

Korrelationen – Signifikanztest

Universität Augsburg University

- Analog zum t-Test mit einem Unterschied:
  - Der Stichprobenkennwert t wird mit Hilfe der Korrelation zweier Stichproben und nicht mit der Differenz der Mittelwerte berechnet.

- **Nullhypothese:**
  - Die empirisch ermittelte Korrelation r zweier Variablen stammt aus einer Grundgesamtheit, in der zwischen den beiden Variablen **keine** Korrelation besteht.

- **Alternativhypothese:**
  - Die empirisch ermittelte Korrelation r zweier Variablen stammt aus einer Grundgesamtheit, in der zwischen den beiden Variablen **eine** Korrelation besteht.

Korrelationen – Signifikanztest

Universität Augsburg University

Ablauf:

1. Berechnung des t-Werts

$$t_{df} = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \quad mit \quad df = N-2$$

2. Teste für das vorher festgelegte Fehlerniveau α gegen einen kritischen t-Wert.
3. Übertrifft der empirische t-Wert diese Grenzmarke, so ist die Korrelation statistisch signifikant.

Korrelationen –
Effekt

Konventionen für die Beurteilung des Effekts

- Wertebereich einer Korrelation:
  - r=-1 perfekt negativer Effekt
  - r=+1 perfekt positiver Effekt
- Nullkorrelation bei r=0

- Bewertung des Effekts abhängig von der jeweiligen Anwendung
- Richtwerte:
  - Kleiner Effekt: |r|=0,10
  - Mittlerer Effekt: |r|=0,30
  - Großer Effekt: |r|=0,50

Zusammenfassung - Wann chi^2 und wann Korrelation?

- Beide Tests untersuchen den Zusammenhang zweier Variablen.
- Unterscheiden sich in der Art und Weise:
  - chi$^{2:}$: Häufigkeiten für Kategorien
  - Korrelation: Paare von intervallskalierten Werten

- Beispiel: Zusammenhang zwischen Alter und Einstellung gegenüber Computerspielen
  - chi$^{2:}$:
    - Alterskategorien: Kinder, Jugendliche, Erwachsene
    - Einstellung: Befürwortung, Ablehnung
  - Korrelation: Paare von intervallskalierten Werten
    - Tatsächliches Alter
    - Bewertung der Zustimmung (z.B. Likert-Skala)