

KAPITEL 2: DIE SUCHE NACH GESCHLECHTSUNTERSCHIEDEN IN KOGNITIVEN FÄHIGKEITEN

A) ARTEN VON WISSENSCHAFTLICHEN UNTERSUCHUNGEN:

1. Beweise und Ergebnisse auf anekdotischem Niveau

Die meisten Menschen haben sehr starke Überzeugungen hinsichtlich Geschlechtsunterschieden (z.B. Frauen sind schlecht im Rechnen). Als Beweis werden dann z.B. eigene Kinder angeführt, die Mathematik hassen, ABER: hier erfolgt Sprung von Mathematikleistungen auf emotionale Ebene!

~~„Ich kenne eine Person, die...“-Phänomen:~~

d.h. **eigene Erfahrungen werden auf Allgemeinheit übertragen**, z.B. Mein Großvater hat geraucht und wurde 100, also kann Rauchen nicht schädlich sein.

Schlecht, weil:

- eigene Erfahrungen müssen **nicht repräsentativ** für Allgemeinheit sein (Einzelfälle!)
- Bias aufgrund von **eigener Erinnerung, eigener Erwartung** (Stereotypen sind SEHR schwierig zu verändern, weil wir dafür Infos aus eigener Umwelt auswählen und uns merken, die mit unseren Überzeugungen übereinstimmen)
- **fehlende Genauigkeit** (z.B. Bruder bekam bessere Noten in Mathematik als Schwester, ABER Schwester weiß nur, dass er besser war, jedoch nicht um wie viele Notenstufen!)

Merke: nur systematische Untersuchung von ausreichend großer Stichprobe, die repräsentativ für die Population ist, erlaubt allgemeine Beschreibung von menschlichem VH.

Mensch hat Tendenz zur Ablehnung von Ergebnissen, die nicht zu seinen Überzeugungen passen.

~~Fragebogen-Daten:~~ meist nicht sehr reliabel, da Self-Reports
→ Gefahr, dass Personen **sozial erwünschte Antworten** geben (z.B. Frauen geben an, dass sie gern Gedichte schreiben, Männer nicht, weil das als unmännlich gelten könnte...)

aus **Anzahl von Personen mit naturwissenschaftlichem Beruf**

→ **keine Rückschlüsse auf Geschlechtsunterschiede daraus möglich!** (Frauen haben es schwieriger, in naturwissenschaftlichen Beruf zu kommen, d.h. aber nicht, dass sie dafür unbegabter sind!)

Fallstudien: stützen sich auf Einzel-Beispiele! **NICHT repräsentativ** für Population!
ABER: Fallstudien gut geeignet als Ausgangspunkt für nähere Erforschung einer Fragestellung

2. Korrelativer Approach mit nicht-randomisierter Personenzuordnung:

- Korrelationsstudien untersuchen **Zusammenhänge zwischen Variablen**, ABER: **keine Kausalbegründungen möglich!**
- Problem der **Scheinkorrelationen** (z.B. Ansteigen des Kunstdüngerverbrauchs und der Katzengeburten im Frühjahr -> Variable dahinter = Jahreszeit!)
- Vor allem in Forschung zu Geschlechtsunterschieden ist **Randomisierung oft nicht möglich**, weil Männer halt in Männergruppe und Frauen in Frauengruppe kommen.
- **Alle Forschungsergebnisse sind Wahrscheinlichkeitsaussagen** (d.h. gefundene Geschlechtsunterschiede könnten auch auf Zufall beruhen...)

3. Beobachtungstechniken:

= **Beobachtung von menschlichem VH in realer Umgebung**

BEISPIEL: Fragestellung: **Sind Buben aggressiver als Mädchen?**
dazu **Beobachtung von Kindern auf Spielplatz**

Vorteil: **direkte Beobachtung von „richtigem“ VH**
(nicht indirekt wie bei Befragung)

Nachteil:

- **Beobachtung ist NIE wirklich objektiv**, weil wir das sehen, was wir sehen wollen... (z.B. wenn Mädchen anderes niederstößt, wird das als weniger aggressiv bewertet als bei Buben)
- Kinder reagieren anders, wenn sie wissen, dass sie beobachtet werden -> **Beobachtung verändert das VH!**
- Ort der Beobachtung spielt eine Rolle (z.B. Geschlechtsunterschiede treten auf Sportplatz auf, nicht aber in Klasse) -> **VH ist kontextabhängig!**

4. Experiment und Quasiexperiment:

- Experiment:**
- Forscher hat **hohen Grad an Kontrolle über die Variablen**, weil
 - **Personen per Zufall der VG oder der KG zugewiesen** werden.
 - Nur bei richtigem Experiment sind **Kausalschlüsse zulässig!**

In Geschlechtsunterschiedforschung eher selten
→ aufgrund von Geschlechtszugehörigkeit
Randomisierung nicht möglich!

Quasi-Experiment: **keine Randomisierung**, sonst wie Experiment
(Manipulation einer UV und Beobachtung der
Auswirkung auf die AV)

wichtig: bei Geschlechtsunterschieds-Forschung spielt **IMMER die jeweilige Umwelt der Personen** bzw. ihre **Lebensgeschichte** mit herein!

Methodischer Würfel von Hendrick, Marvel & Barrington (1990):

3 Dimensionen:

- | | |
|----------------------------------|---|
| 1) Untersuchungsdesign: | deskriptive, korrelative,
experimentelle
und quasi-experimentelle Designs
liefern unterschiedliche Arten von
Infos! |
| 2) Untersuchungs-Setting: | Labor vs. Feld |
| 3) Datenerhebung: | Beobachtung vs. Self-Report |

5. Faktorenanalytische Zugänge:

Kognition ist kein einzelnes, homogenes Konzept
-> es gibt **mehrere kognitive Fähigkeiten**, Person kann in einer,
mehreren oder allen begabt/unbegabt sein

Eine der wichtigsten Unterscheidungen ist diejenige zwischen:

- o **verbalen Fähigkeiten** und
- o **räumlichen Fähigkeiten**

BEISPIEL: Um **Geschlechtsunterschiede FA zu untersuchen**
 → **große Zahl von Männern und Frauen** und **mehrere Tests** von verbalen Fähigkeiten (z.B. verbale Analogien, Wortschatz) und räumlichen Fähigkeiten (z.B. Landkarte, Puzzle) notwendig)

Wenn diese 4 Tests wirklich 2 verschiedene Fähigkeiten prüfen, dann ergibt FA der Ergebnisse 2 Faktoren:

- einen für verbale Fähigkeiten (hier hohe Ladungen für Scores der beiden verbalen Tests)
- einen für räumliche Fähigkeiten (hier hohe Ladungen für Scores der beiden räumlichen Tests)

Ergebnis:

- Ergibt FA bei beiden Gruppen (d.h. Frauen und Männern) die gleichen Faktoren, dann Schluss zulässig, dass Frauen und Männer ähnliche FA-Strukturen aufweisen
- Ergibt FA für Frauen nur 1 Faktor -> bei Frauen liegt nur EINE Dimension zugrunde [ABER: es könnte auch sein, dass die 4 Tests einfach nur dasselbe messen...]
- 2 Faktoren bei Männern, aber nur 1 Faktor bei Frauen -> es gibt Geschlechtsunterschiede

6. Validität, Fairness und Bias:

Im Allgemeinen haben Tests 2 Zwecke:

- Bewertung von Gelerntem
- Voraussagen für Zukunft

prädikative Validität: ist der Test in der Lage, Voraussagen über die Zukunft möglich zu machen

Messinstrument = fair (bzw. biasfrei), wenn es Erfolg in Kriteriumsvariable für unterschiedliche Personengruppen gleich gut vorhersagen kann

ABER: **genauso wichtig** sind die **sozialen Konsequenzen, die Testergebnis für Person nach sich zieht** (z.B. Personalauslese)

B) DIE WISSENSCHAFTLICHEN ERGEBNISSE VERSTEHEN:1. Die Logik des Hypothesentestens:

H0: Es gibt keine Unterschiede in untersuchter Variable zwischen Frauen und Männern, Unterschiede in den Ergebnissen sind rein zufällig.

H1: Es gibt systematische Unterschiede in den untersuchten Variablen zwischen Frauen und Männern, Unterschiede sind überzufällig.

⇒ statistischer Test überprüft, wie wahrscheinlich das beobachtete Ergebnis ist. Ergeben sich signifikante [Mittelwerts-]Unterschiede zwischen den untersuchten Gruppen, dann Verwerfung der H0, ABER: nur mit bestimmter Wahrscheinlichkeit!!!

2. Das Problem nicht-signifikanter Ergebnisse:

Wenn Ergebnis **statistisch nicht signifikant** ist, heißt das **NICHT automatisch, dass es den untersuchten Geschlechtsunterschied nicht gibt**

→ H0 wird ja immer nur auf bestimmtem Niveau verworfen (z.B. Alpha = 5%, d.h. in 5 % der Fälle könnte ich mich irren = Alpha-Fehler!)

BEISPIEL: Unterschiede zwischen Frauen und Männern im Buchstabieren
→ Ergebnis ist nicht statistisch signifikant

Das könnte heißen:

- Es gibt **keine Unterschiede** zwischen Frauen und Männern
- Es gibt **Unterschiede**, aber sie wurden **nicht gefunden** (vgl. Alpha-Fehler -> bei $\alpha = 5\%$ irre ich mich in 100 Fällen 5x)
- **Test war zu leicht oder zu schwer**, daher traten Unterschiede nicht auf (zu wenig Varianzunterschiede in den beiden Gruppen)

Merke: NIE sagen, die H0 (oder die H1) ist wahr oder falsch
→ **nur Wahrscheinlichkeitsaussagen sind möglich**, weil es unmöglich ist, eine Hypothese zu verifizieren (dafür müsste ich ALLE Elemente einer Population untersuchen, was meist nicht möglich ist!)

Signifikante Studien haben [aber meiner Meinung nach NUR in den Augen der scientific community!] **mehr Gewicht** als nicht-signifikante.

Grund: nicht-signifikante Studien könnten Mängel aufweisen, wodurch Unterschiede nicht aufgefunden werden konnten, wie z.B. zu wenige VPn, schlechte Messinstrumente, schlechtes Design, etc.

Rosenthal & Rubin (1985): Es ist **unmöglich, Aussagen über ein Phänomen** zu machen, wenn dessen **Untersuchung nur in einer einzigen Studie** erfolgte

BEISPIEL: **mehrere Studien zu Geschlechtsunterschied in bestimmter Variable**

- $\frac{1}{2}$ **der Studien findet einen Geschlechtsunterschied**, $\frac{1}{2}$ der Studien aber nicht -> **keine Aussagen über Existenz dieses Geschlechtsunterschieds möglich!**
- **alle Studien finden Geschlechtsunterschied** -> es ist wahrscheinlich, dass es diesen tatsächlich gibt, ABER:
 - Ergebnisse könnten auch auf Besonderheiten der verwendeten Messinstrumente zurückzuführen sein
 - Ergebnisse könnten nur in der untersuchten Subpopulation (z.B. AHS-SchülerInnen) vorkommen

wichtig: **VOR Datenerhebung festlegen, ab wann man von einem Geschlechtsunterschied sprechen kann**, z.B. wenn sich Frauen und Männer in Intelligenztest im Mittel um 2 Wertpunkte unterscheiden

Merke: Auch wenn **Ergebnis der Hypothesenprüfung signifikant** ist, **heißt das noch lange nicht, dass es auch bedeutsam ist** [-> Effektstärke berechnen!]

3. Einfluss auf die Ergebnisse:

In Geschlechtsunterschied-Forschung gibt es mehrere Fallen im Zusammenhang mit den Stichproben:

a) Zusammensetzung der Stichproben:

Geschlechtsunterschiedforschung möchte Unterschiede zwischen Frauen und Männern nachweisen, dafür Auswahl von Stichproben aus Population,

ABER:

- **Problem der Selbst-Selektion**, d.h. wenn ich z.B. mathematische Fähigkeitsunterschiede zwischen Mädchen und Burschen untersuchen möchte und dafür an Realgymnasien testen gehe, habe ich sicher andere Ergebnisse, als wenn ich in Hauptschulen gehe...
- Bei Untersuchung z.B. in Schulklassen **keine Randomisierung möglich**, daher auch **keine Kausalbegründungen**

- **Geschlechtsunterschiede** z.B. in Schulklassen **können mit vielem zusammenhängen**, z.B. unterschiedliche Schulen, unterschiedliche Kinder mit unterschiedlichem Background, Motivation, etc. unterschiedliche Lehrer, usw.

Abhilfe hier durch partielle Korrelation (-> Herauspartialisieren von Störvariable) **oder Kovarianzanalyse** (-> Konstanthalten von Störvariable),

ABER:

- Bias krieg ich trotzdem nie ganz heraus und
- außerdem müssen für Verfahren auch bestimmte Voraussetzungen erfüllt sein, wie z.B. homogene Varianzen, NV, lineare Effekte in allen Gruppen, usw. (die es aber oft nicht sind...)

Merke: **Ergebnisse hängen immer auch von der Stichprobe ab** (z.B. Bildungsstand, Alter, sozioökonomischer Status, Motivation, usw.) -> muss bei Interpretation der Ergebnisse bedacht werden!

BEISPIEL: Untersuchung, ob Frauen in unterschiedlichen Stadien des Menstruationszyklus Unterschiede in kognitiven Leistungen aufweisen

Frage: **Wen nehme ich als KG?**

(Frauen nach Wechsel, Frauen nach Hysterektomie?, Männer????)

→ wurscht, wen Ergebnisse werden sich je nach KG unterscheiden...

b) Stichprobengröße:

Große Stichproben erlauben gute Schätzung von Populationsparametern (d.h. den wahren Werten in der Population)

wichtig dabei: Höhe der Varianz in der Population

- **wenig Varianz:** kleine Stichprobe genügt
- **große Varianz:** größere Stichprobe notwendig

In Geschlechtsunterschieds-Literatur heißt es, dass **bei Männern generell höhere Streubreiten** auftreten

→ wenn das stimmte, dann müssten Männer-Stichproben immer größer sein als Frauenstichproben (wird aber in der Praxis nirgends so gemacht...)

Große Stichprobe -> Tendenz, dass Ergebnis statistische signifikant wird, obwohl in Wirklichkeit Geschlechtsunterschied gar nicht existiert, z.B.: Viele Studien mit kleinen Stichproben ergeben keine Geschlechtsunterschiede in bestimmter Variable, Studien mit großen Stichproben zur selben Variable dagegen schon

Daher: **Effektstärke festlegen und danach ideale Stichprobengröße wählen!**

c) ungerechtfertigte Generalisierungen:

- In Geschlechtsunterschieds-Forschung **Ziehung von Stichproben oft aus „abnormen“ Populationen** (z.B. Personen mit extrem hohen Hormonkonzentrationen und deren Auswirkung auf bestimmtes Personenmerkmal), mit den Ergebnissen dann aber Aussagen über „normale“ Population -> im Grunde NICHT zulässig, da nicht verallgemeinerbar! (Hormonkonzentration könnte sich auf andere, nicht untersuchte Variable auswirken und DIE ist dann verantwortlich für den Geschlechtsunterschied...)
- **Untersuchung von Geschlechtsunterschieden an Tieren** (z.B. Ratten, Affen) -> Übertragung der Ergebnisse auf den Menschen
ABER: Hormone spielen z.B. bei Tieren wesentlich bedeutendere Rolle als bei Menschen, bei Menschen wiederum bestimmen kognitive und soziale Variablen stärker das VH als bei Tieren.

BEISPIEL: Männliche Kanarienvögel entwickeln komplexes Sing-Verhalten aufgrund von Testosteron, d.h. aber NICHT, dass Testosteron für menschliches Sprach-VH bedeutsam ist (-> Unterschiede in den Gehirnen!)

d) Wechselwirkungen Geschlecht x Alter:

Einige Geschlechtsunterschiede verändern sich mit zunehmendem Lebensalter (z.B. treten in Pubertät zu Tage, nehmen im Alter ab), daher: bei Untersuchung von Geschlechtsunterschieden NICHT nur z.B. Studenten als VPn nehmen!

Aufgrund von Alter Auftreten von Kohorteneffekten möglich!

BEISPIEL: Unterschiede in räumlicher Orientierung beim Autofahren zwischen 75-jährigen Frauen und 75-jährigen Männern; Ergebnis: Frauen können schlechter Landkarten lesen...

ABER: Gründe dafür könnten auch sein:

- Effekte systematischer Bildungsunterschiede
- Effekte der geschlechtsspezifischen Lebenserfahrungen
- Effekte des Alters

Mehrzahl der Studien verwendet Studenten als VPN

➔ Mehrheit der Bevölkerung ist aber NICHT Akademiker

➔ Mehrheit der Bevölkerung ist älter als Studenten

daher Frage: Inwiefern sind hier gefundene Geschlechtsunterschiede überhaupt generalisierbar?

e) Replizierbarkeit von Ergebnissen:

„gute“ Forschung bedeutet, das man **Ergebnisse einer Studie an neuen Stichproben überprüft**

- erst wenn Ergebnisse der Original-Studie derart bestätigt werden konnten, kann man annehmen, dass die gefundenen Geschlechtsunterschiede wahrscheinlich geben wird!
ABER: **In der Praxis ist eine solche Vorgangsweise sehr selten...**

f) Messung:

Art der Messung hat direkten Einfluss auf die Ergebnisse!

Geschlecht wird üblicherweise als dichotome Variable

(weiblich/männlich) erhoben,

ABER: Manchmal wäre es günstig, den Grad der Männlichkeit/Weiblichkeit zu messen (z.B. bei Fragen im Zusammenhang mit Untersuchung von Hormonen, Chromosomen, Selbstbild, etc.)

Problem:

- Es ist **nicht immer klar, WAS Forscher mit „Geschlecht“ meinen**, vgl. biologisches Geschlecht, Geschlechtsidentität, sexuelle Präferenz, usw. -> müssen NICHT identisch sein! (vgl. sex vs. gender)
- **Frage der Geschlechtsrolle**: Frauen/Männer können hier unterschiedliche weibliche/männliche Anteile aufweisen. Für Frage, ob sich Person gemäß herrschenden Stereotypen für männliches/weibliches VH verhält (d.h. für qualitative Geschlechtsunterschiede), wäre Messung auf Intervall- oder Analogskala oft günstiger

g) Publikationszeitpunkt und Geschlecht des Forschers:

Publikationszeitpunkt:

Studien aus 40er Jahren berichten über größere Geschlechtsunterschiede als neuere Studien, z.B. in der Lesefähigkeit von Studienanfängern;

Grund: vor 60 Jahren gab es wesentlich weniger weibliche als männliche Studienanfänger. Weibliche Studienanfänger waren besonders intelligent und kamen aus gehobenen Verhältnissen -> DAS könnte der Grund für die größeren Geschlechtsunterschiede sein...

Geschlecht des Forschers:

nimmt auch **(auch unbewusst!) Einfluss auf die Ergebnisse**,

BEISPIEL: Forscherin ist Feministin, Forscher nicht, beide untersuchen dieselbe Fragestellung -> kommen zu unterschiedlichen Ergebnissen!

persönliche Erfahrungen und Einstellungen des Forschers/der Forscherin können **Bias in den Ergebnissen** bewirken

BEISPIEL: Hyde & Slinn (1988): Forscherinnen finden systematisch öfter verbale Überlegenheit von Frauen als Forscher

h) Moderator- und Situationsvariablen:

- **Situation, in der Daten erhoben werden, kann Ergebnisse beeinflussen**

BEISPIEL: Eagly (1987): In Laborsituation werden weniger Geschlechtsunterschiede gefunden als in Feldsituation.

- **Grund: Verhalten ist IMMER kontextgebunden!**

BEISPIEL: Untersuchung von Geschlechtsunterschieden beim Anbieten von Hilfe; erfolgt meist mit Frage, ob man einem Fremden in einer gefährlichen Situation helfen würde -> Männer in der Regel höhere Werte als Frauen...
ABER: mit SO einer Frage kriegt man tatsächliche Geschlechtsunterschiede NICHT heraus, denn Frauen bieten z.B. in Familie und Partnerschaft wesentlich öfter und mehr Hilfe an als Männer...

- **Laborexperimente** bedeuten oft Vorliegen einer **artificialen Situation**, daher **andere Ergebnisse als bei Feldexperimenten!**
- **Personen verhalten sich anders, wenn sie wissen, dass sie „beobachtet“ werden.**

BEISPIEL: - Mann gibt öffentlich nicht zu, dass er privat gern strickt, weil Gesellschaft das für feminines Verhalten ansieht...

- Mann verhält sich bei weiblichen VL anders als bei männlichem VL (und umgekehrt!)

i) Multivariate Indikatoren:

Messung der Variable Geschlecht erfolgt normalerweise als **Self-Report** (d.h. Person kreuzt an, ob sie weiblich oder männlich ist)

Messung von kognitiven Fähigkeiten (z.B. räumliche Fähigkeiten) dagegen **auf vielfältige Art mit vielfältigen Aufgabenstellungen** (z.B. Zurechtfinden in Wald ohne Landmarks, in Stadt mit Landmarks, Unterschiede zwischen 2 Gebäuden herausfinden, usw.)

→ Frage: **Inwiefern messen diese Items aus verschiedenen Tests tatsächlich dasselbe bzw. nur EINE einzige Fähigkeit?**
[vgl. Würfelaufgaben, die auch logisches Schließen mitprüfen!]

daher: mehrere Tests für räumliche Fähigkeiten vorgeben, wenn Geschlechtsunterschiede gefunden werden sollen:

- findet man sie z.B. in 4 Tests, dann höhere Wahrscheinlichkeit, dass es sie tatsächlich gibt, als wenn man nur 1 Test einsetzt
- findet man sie z.B. in 3 Tests, nicht aber im 4., dann überlegen, inwiefern der 4. Test wirklich DAS misst, das er zu messen vorgibt!

ABER: **wenn man multiple Indikatoren untersucht, dann muss man auch multivariat auswerten** (und nicht eine ANOVA für jeden einzelnen Test getrennt rechnen...)

4. Statistische und praktische Signifikanz:

Untersuchungsergebnisse sind **IMMER Wahrscheinlichkeitsresultate**, d.h. bei $\alpha = 5\%$ irren 5 von 100 Studien.

Wenn **Ergebnis statistisch signifikant** ist, kann man schließen dass Geschlechtsunterschied wahrscheinlich ist, wenn

- Ergebnis theoretisch untermauert ist
- Ergebnis mit verschiedenen Stichproben repliziert werden konnte

ABER: **Unterschied zwischen Vergleichsgruppen muss nicht nur signifikant, sondern auch bedeutsam sein** (d.h. es muss praktische Signifikanz haben)

BEISPIEL: Mädchen und Burschen unterscheiden sich signifikant in der Zeit, die sie mit Fernsehen verbringen

- 0.4 Stunden / Woche = statistisch signifikant
- d.h. die Unterschiede betragen pro Tag 3.4 Minuten... Das ist praktisch NICHT bedeutsam!

a) Variabilität und Verteilungsform:

Um Geschlechtsunterschied in bestimmter Variable zu verstehen, muss man sich auch die Form der Verteilung in Frauen- und Männergruppe anschauen -> wichtig zur Interpretation der Ergebnisse!

4 Möglichkeiten (bei 2 Vergleichsgruppen Frauen vs. Männer):

1) M. und SD. = gleich:

Verteilung der Scores für bestimmte Variable weist für Frauen und Männer identische Kurvenform auf -> keinerlei Unterschiede

2) M. = gleich, SD. = unterschiedlich:

z.B. Männer-Kurve ist steil, Frauen-Kurve ist flach, d.h.

- Werte der Männer liegen alle in der Nähe von M, sehr homogen
- Werte der Frauen haben höheren Streubereich von sehr schlecht bis sehr gut, sehr inhomogen

3) M. = unterschiedlich, SD. = gleich:

hierbei wichtig: Überschneidungen der 2 Kurven:

- **keine Überschneidung**, Kurven liegen weit auseinander -> hohe M-Unterschiede zwischen den Gruppen
- **Überschneidung**, Kurven liegen mehr oder weniger eng beieinander -> Signifikanzprüfung zeigt, ob und inwiefern M-Unterschiede signifikant sind [bei SD = gleich -> homogene Varianzen, daher Prüfung mit t-Test, wenn 2 Gruppen]

4) M. = unterschiedlich, SD. = unterschiedlich:

Gruppen unterscheiden sich nicht nur hinsichtlich des M, sondern auch in Verteilung der Werte [-> inhomogene Varianzen, daher Prüfung mit U-Test]

b) Mittelwert und Gipfel der Verteilung:1) @ Gipfel der Verteilung:

BEISPIELE:

- M = gleich, Verteilung ist **linksgipfelig** (d.h. Schwanz zeigt nach rechts), d.h. es gibt bestimmten Anteil von VPn, deren Werte wesentlich höher liegen als M
- M = gleich, Verteilung ist **rechtsgipfelig** (d.h. Schwanz zeigt nach links), d.h. es gibt einen bestimmten Anteil von VPn, deren Werte wesentlich niedriger sind als M

2) **Konfidenzintervall für M:** [Voraussetzung = Normalverteilung!]

aus KI für M kann man schließen, wie sich M (Stichprobenmittelwert) zu μ (Populationsmittelwert) verhält. KI mit 95 % heißt: mit 95 %-iger Wahrscheinlichkeit liegt μ innerhalb des KI

- KI sehr schmal, d.h. meine Werte passen ziemlich gut zu μ
- KI sehr breit, d.h. meine Werte sind in Hinblick auf μ noch ziemlich unverlässlich, größere Stichprobe notwendig

3) **Tail-Ratio** [= Verhältnis zum Schwanz der Verteilung]

BEISPIEL:

Anzahl Männer in oberen 5 % der Stichprobe	39	
-----	----	= .52
Anzahl Männer in gesamter Stichprobe	750	
-----	-----	= 1.18
Anzahl Frauen in oberen 5 % der Stichprobe	11	
-----	----	= .44
Anzahl Frauen in gesamter Stichprobe	250	

d.h. für jede Frau in den oberen 5 % der Stichprobe findet man 1.18 Männer

- Tail-Ratio wird verwendet **zu Selektionszwecken**, z.B. Zulassung zu Studium (nur wer oberhalb eines bestimmten Scores liegt, wird von Uni aufgenommen)
- wichtig z.B. bei **Evaluation von 2 Programmen** (welches ist erfolgsversprechender als das andere?)
- vor allem dann verwenden, **wenn M-Unterschiede gering sind, Verteilungen aber Schwänze haben**

BEISPIEL: Ergebnisse von bestimmtem Test:

- Schwanz nach links, M eher rechts -> Test war generell leicht, aber einige haben schlechte Ergebnisse
- Schwanz nach recht, M eher links -> Test war generell schwer, aber einige haben gute Ergebnisse

c) **Meta-Analysen:**

Erster Versuch = **Studie von Maccoby & Jacklin (1974)** über alle Studien zu Geschlechtsunterschieden, die 1964-1974 in amerikanischen Journals veröffentlicht wurden -> nur wenn Geschlechtsunterschied in bestimmter Variable in sehr vielen Studien gefunden wurde, dann Annahme, dass es ihn tatsächlich geben könnte.

Meta-Analyse = **Analyse von Analysen**, d.h. gemeinsame Analyse von vielen einzelnen Untersuchungsergebnissen

vor allem wichtig **in Gebieten, in denen sehr viel geforscht und publiziert wird**

gibt nicht nur **Anzahl der Studien** zu Geschlechtsunterschied in bestimmter Variable an, sondern auch **wie hoch der gefundene Unterschied** ist

File-Drawer-Problem: [= Retrieval Bias]

Einzelne Studien, die in Meta-Analyse untersucht werden, müssen repräsentativ für die Forschung auf diesem Gebiet sein.

Repräsentative Ergebnisse findet man vor allem in [peer-reviewed] Journals.

ABER: **Bias, weil in erster Linie Studien publiziert werden, die signifikante Ergebnisse haben**, d.h. 95% Studien landen in Schublade, da nicht signifikant, 5% veröffentlicht, da signifikant [aber auch hier Alpha-Fehler...], d.h. File-Drawer-Problem bzw. Publikationsbias

Wo findet man nicht-publizierte Studien?

- **Dissertation Abstracts** (aber hier Gefahr von Bias, weil Verfasser ja unbedingt signifikantes Ergebnis haben wollte)
- z.B. in **ERIC** (pädagogische Datenbank; hier auch Kongressbeiträge u.a. nicht-veröffentlichte Präsentationen)

In den letzten Jahren nimmt Tendenz, nur signifikante Ergebnisse zu publizieren gottlob eh ab, ABER: nicht-signifikante Ergebnisse werden dennoch vielfach mit scheelen Augen betrachtet, weil Nicht-Signifikanz auch auf Fehlern in Design und Methodik basieren könnte...

Abhilfe gegen File-Drawer-Problem bei Meta-Analyse:

➔ Berechnung der Anzahl nicht-veröffentlichter Studien = **Fail-Safe-Zahl von Rosenthal:**

Formel dazu: $5k + 10$

- k = Anzahl der Studien, die in Meta-Analyse untersucht werden;
- 5 weil Rosenthal meint, dass es unwahrscheinlich ist, dass die Anzahl nicht signifikanter Studien mehr als 5x so hoch ist als die Anzahl signifikanter Studien;
- 10 = Sicherheitszuschlag

d) **Effektstärken:**

Dafür gibt es verschiedene Maße:

1) ω^2 **(Omega-Quadrat):**

= Maß für den **Anteil der Gesamtvarianz, der durch eine bestimmte Variable erklärt wird** (z.B. Geschlecht der VPn)

BEISPIEL: **Reaktionsgeschwindigkeit auf bestimmten vorgegebenen Reiz**

4 VG: Frauen jung, Frauen alt, Männer jung, Männer alt

Je nach Art des vorgegebenen Reizes können Geschlechtsunterschiede auftreten

Annahmen:

- Ältere haben durchgehend längere Reaktionszeiten
➔ in diesem Fall wird ω^2 für die Variable Alter hoch sein
- Männer haben durchgehend kürzere Reaktionszeiten
➔ in diesem Fall wird ω^2 für die Variable Geschlecht hoch sein.

⇒ Vergleicht man nun die beiden ω^2 für Alter und Geschlecht, kann man sagen, welche der beiden Variablen sich stärker auf die Reaktionszeiten auswirkt

In **Meta-Analyse** wird ω^2 für jedes Ergebnis der Studien, in denen dieselbe Fähigkeit gemessen wurde, berechnet

➔ dann daraus **Berechnung eines mittleren ω^2 für alle Studien**, d.h. Ergebnisse der einzelnen Untersuchungen zu bestimmtem Thema werden so in mittlerer Effektstärke zusammengefasst

ABER: **Viele Studien geben Effektstärken gar nicht an...**
(Grund: Effekte sind oft sehr klein, d.h. untersuchte Variable ist nicht besonders entscheidend für die untersuchte Fähigkeit)

Interpretation von ω^2 (bzw. jedem Verhältnis von erklärter Varianz) ist dadurch **beschränkt**, weil Wert immer von den untersuchten Variablen abhängt

→ z.B. ω^2 für Geschlecht aus Untersuchung, die Geschlecht und Alter untersucht, kann nicht mit ω^2 für Geschlecht aus Untersuchung, die Geschlecht und sozioökonomische Variable untersucht, verglichen werden!

⇒ **Vergleich von ω^2 in mehreren Studien geht nur, wenn exakt dieselben Variablen untersucht wurden!**

2) **Effektstärke d** (Cohen, 1977)

Formel:
$$d = \frac{M_{(\text{Männer})} - M_{(\text{Frauen})}}{SD}$$

d.h. gemessen wird der **Unterschied der Differenz zwischen den M der beiden Vergleichsgruppen** (Männer vs. Frauen), **bezogen auf die SD in der Gesamtstichprobe**

Bei Normalverteilung liegen die Scores üblicherweise in folgenden Bereichen:

- 32 % der Scores im Bereich $M \pm 1 \text{ SD}$
- 16 % der Scores im Bereich zwischen 1 SD und 2 SD
- 2% der Scores im Bereich zwischen 2 SD und 3 SD

ABER: Für Berechnung von d müssen die **SD der beiden Vergleichsgruppen homogen** sein!

Es gilt:

- **d = groß:** wenn **M-Unterschiede groß** und **Varianzen innerhalb klein**
- **d = klein:** wenn **M-Unterschiede klein** und **Varianzen innerhalb groß**

Unterschied zu ω^2 : **d gibt auch die Richtung des Effekts an**, d.h. (siehe Formel oben!)

- **positiver Wert:** Männer haben höhere Scores
- **negativer Wert:** Frauen haben höhere Scores

Insgesamt:

- **hohes d** -> **große Geschlechtsunterschiede**
- **niedriges d** -> **kleine Geschlechtsunterschiede**

Grund für Bezugnahme auf SD:

wenn 2 Tests unterschiedliche Wertebereiche haben (z.B. SAT von 200-800 und IQ-Test-M = 100), können Ergebnisse in beiden Tests verglichen werden, wenn man auf Gesamt-SD bezieht

Klassifikation der Effektstärke d nach Cohen:

- $d = .2$ = kleiner Effekt
- $d = .5$ = mittlerer Effekt
- $d = .8$ = großer Effekt

ABER: Gefahr der Missinterpretation, denn z.B. in Medizin kann ein Winz-Effekt sehr bedeutsam, da lebensrettend sein! D.h. man muss **berücksichtigen, WAS untersucht wurde!**

In Meta-Analyse Durchführung von Homogenitätstests für Effektstärken unterschiedlicher Studien:

Annahme:

$\frac{1}{2}$ der Studien zeigt großen Effekt zugunsten der Männer (d = positiv und hoch), $\frac{1}{2}$ der Studien zeigt großen Effekt zugunsten der Frauen (d = negativ und hoch) -> Schluss liegt nahe, dass es keinen Geschlechtsunterschied gibt...

ABER: Ergibt Homogenitätstest **Inhomogenität der Ergebnisse**, dann heißt das, dass **Stichproben aus unterschiedlichen Populationen stammen** (Unterschiede gehen daher NICHT auf Geschlecht zurück, sondern auf andere Moderator-Variablen). In diesem Fall **Teilung der Studien in Subkategorien**

BEISPIEL: Hyde (1986): **Meta-Analyse über Geschlechtsunterschiede hinsichtlich Aggression**
 → Homogenitätstest ergab Inhomogenität, daher Unterkategorie Alter
 → Ergebnis: Geschlechtsunterschiede hinsichtlich Aggression sind bei Jungen höher als bei Alten

3) Binomiale Effektstärke (Rosenthal & Rubin, 1982):

- Da in Geschlechtsunterschieds-Forschung **immer 2 Gruppen** (Frauen vs. Männer) verglichen werden binomiale Effektstärke berechnen
- = **Prozentsatz für jedes Geschlecht, der über dem Mittel der Gesamtstichprobe liegt**
- Halpern hält das für gutes Maß, wird aber in Forschung nur **selten verwendet**

4) Wann ist ein Effekt groß?

Auch wenn Effektstärke ω^2 oder d angegeben ist, muss Leser der Studie selbst entscheiden, ob der gefundene Effekt groß genug ist, um bedeutsam zu sein -> **Entscheidung hängt davon ab, WAS untersucht wurde**

BEISPIEL: In Medizin können extrem kleine Effekte extrem große Bedeutung haben (z.B. wenn Medikament erfunden wird, das Krebs bei sehr wenigen Menschen heilt)

5) Vergleich der Effektstärken-Maße:

wichtig sind 3 Dinge:

- ist Ergebnis signifikant?
- ist Effektstärke groß genug, um theoretisch bedeutsam zu sein?
- ist Effektstärke eines signifikanten Ergebnisses groß genug, um praktisch bedeutsam zu sein?

BEISPIEL: Eagly (1987): **Vergleich von Effektgrößen für aggressives VH**

- ω^2 : nur 2 % der Gesamtvarianz von aggressivem VH wird durch Geschlecht erklärt
 - $d = .29$, d.h. kleiner Effekt (ca. 1/3 SD Unterschied)
 - binomiale Effektstärke: 43 % der Frauen und 57 % der Männer lagen über dem für aggressives VH gemessenen M
- ⇒ allen 3 Maßen liegen dieselben Daten zugrunde, dennoch **Riesenunterschiede, je nachdem WIE Effektstärke berechnet wird!**

D.h. auch wenn **nur wenig % der Gesamtvarianz erklärt** werden, kann binomiale Effektstärke zeigen, dass die **Unterschiede zwischen den Vergleichsgruppen gewaltig** sind. Das sollte man beim Lesen empirischer Ergebnisse berücksichtigen.

e) **Interaktion von Variablen:**

Geschlechtsunterschiede in kognitiven Fähigkeiten müssen nicht auf Geschlecht allein zurückgehen, sondern **komplexes Zusammenspiel von verschiedenen anderen Variablen** wie Alter, Stellung in der Geschwisterreihe, sozioökonomischer Statuts, kultureller Hintergrund, Lernerfahrung, usw.

BEISPIEL: reiche, erstgeborene Frauen haben höhere verbale Fähigkeiten als zweitgeborene Frauen aus Unterschicht

→ Grund: Eltern von Ersteren konnten sich Zeit nehmen, mit Kindern zu sprechen, ihnen vorzulesen, etc.

Merke: **Entwicklung von kognitiven Fähigkeiten** wird (neben Geschlecht) **durch Bündel von Variablen beeinflusst:**

- **soziodemografische** Variablen (z.B. Alter, Wohnort)
- **psychologische** Variablen (z.B. Motivation)
- **biologische** Variablen (z.B. Gesundheitszustand)
- **Lebensgeschichte** (z.B. Ausbildung)

⇒ Höhe der Geschlechtsunterschiede ist auch abhängig von Ausprägungen anderer Variablen

5. Entwicklungsaspekte:

Kognitive Fähigkeiten sind nicht statisch, sondern **verändern sich über die Lebensspanne hinweg**

→ **Einfluss von altersabhängigen biologischen und soziologischen Veränderungen**

D.h. Geschlechtsunterschiede können sich z.B. mit dem Alter verändern

daher: **KEINE Aussagen über Geschlechtsunterschiede möglich, ohne Berücksichtigung von Alter, etc.**

Querschnitt-Studien:

- Stichprobe aus Population zu bestimmtem Zeitpunkt -> **Personen unterschiedlichen Alters** werden **einmal untersucht**
- **Problem dabei = Kohorteneffekte** (d.h. Personen mit gleichem Alter haben ähnliche Lebenserfahrungen aufgrund ihrer Zugehörigkeit zu ihrer Generation)

BEISPIEL: **Fähigkeit zum Kartenlesen** getestet an Frauen und Männern aus 3 Altersgruppen:

- alte Frauen haben vermutlich nie Autofahren gelernt, daher keine Übung im Kartenlesen; ABER: junge Frauen können Auto fahren und daher auch Karten lesen
- bei Männern kein solcher Unterschied, da auch alte Männer zeitlebens Auto gefahren sind

Längsschnitt-Studien:

- Stichprobe von **Personen gleichen Alters** wird zu **mehreren Zeitpunkten untersucht** (als Jugendliche – Erwachsene – Alte)
- **Probleme dabei:**
 - bei Testung in kurzen Intervallen hintereinander kann **vorhergehende Testung die nachfolgenden beeinflussen**
 - **VP-Schwund** durch Tod, Umzug, Verweigerung zur Fortsetzung der Testreihe, usw.
 - es dauert **viele Jahre, bis Ergebnisse vorliegen** (und womöglich sind die gefundenen Ergebnisse dann gar nimmer soooo aktuell...

auch **Kombination** möglich (Personen verschiedenen Alters werden im Abstand von z.B. 5 Jahren untersucht)

6. Selbsterfüllende Prophezeiungen:

Erwartungen des Untersuchers können Ergebnisse beeinflussen
= **Rosenthal-Effekt** (1966)

Experiment von Rosenthal:

Lehrern wurde gesagt, dass bestimmte SchülerInnen hohe Leistungen in IQ-Test erbracht hätten -> diese SchülerInnen bekamen dann tatsächlich bessere Noten. Lehrer berichteten, dass sie besonders interessiert, ausgeglichener, etc. wären als die anderen

ABER: Diese **SchülerInnen waren bei nachfolgender Testung tatsächlich besser** (Grund: Wissen, dass Lehrer mich für besonders tüchtig hält, bewirkt, dass ich tatsächlich besser werde) [könnte aber auch sein, weil sie aufgrund ihrer vermeintlichen hohen Intelligenz von Lehrern anders behandelt wurden...]

Vor allem Geschlechtsunterschieds-Forschung ist anfällig dafür!

- ➔ wenn VL überzeugt ist, dass Frauen in bestimmtem Test bessere Leistungen erbringen als Männer, dann behandelt er (unbewusst!) Frauen anders (z.B. ist freundlicher, ermuntert mehr, lässt länger Zeit, usw.)

→ Erwartung der VPn beeinflusst Testergebnisse

BEISPIEL: Frauen, die überzeugt sind, nicht rechnen zu können, schneiden bei Rechenaufgaben schlechter ab, auch wenn sie objektiv betrachtet gute Rechnerinnen sind

→ es kann auch vorkommen, dass **VPn in Test bewusst falsche Antworten geben** werden (z.B. wenn Frau überzeugt ist, gut in gemessener Fähigkeit zu sein, sei unweiblich...)

Abhilfe:

- **„blinder“ VL**
(d.h. Untersucher weiß nicht, was eigentlich untersucht wird; weiß nicht, welches Geschlecht die VPn haben und kann daher keine geschlechtsbezogenen Erwartungen ausbilden -> geht aber nur, wenn er die VPn nicht sieht bzw. bei gleich angezogenen Kleinkindern)
- **„blinde“ VPn**
(d.h. Personen wissen nicht, dass in Wirklichkeit Geschlechtsunterschiede untersucht werden)