

# PSYC2104 - Kvanitativ metode A:

## Alt du (bank i bordet) skal trenge å kunne til eksamen

*Formålet med dette dokumentet er å være en (fingers crossed) komplett oversikt over alt du må kunne til eksamen. Først kommer kunnskaps- og ferdighetsmålene med kommentarer. Så kommer en liste over alle konsepter/begreper du skal kunne, basert på en gjennomgang av de 4 siste utgavene av eksamen og sensorveiledning. Kan du alle disse kan du alt som har vært nødvendig for å få en A på disse.*

### Kunnskapsmål

- Ha inngående kjennskap til psykologisk testteori
- Kjenne til prinsippene bak operasjonalisering av latente variabler og formative begrep
- Vite hva "klassisk målingsteori" er og gjøre rede for ulike design for estimering av reliabilitet
- Beskrive og diskutere validitet i forbindelse med målinger
- Vite hvordan man kan undersøke uni- og multidimensjonalitet i måleinstrumenter. Her vil bruk av faktoranalyse være sentralt.
- Diskutere kausalitet og ikke-eksperimentell forskning, herunder kunne gjøre rede for begrep som spuriøsitet, supressoreffekter og statistisk kontroll. Medierings- og modereringseffekter er også sentralt
- Beskrive prinsippene bak multippel lineær regresjonsanalyse og gjøre rede for begrepene forklart varians, ustandardiserte og standardiserte regresjonskoeffisienter, signifikanstesting i regresjonsanalyse og forutsetninger for multippel lineær regresjonsanalyse

### Ferdighetsmål

- Estimere reliabilitet til måleskalaer med ulike design
- Kunne beregne standardfeil til målinger
- Kunne beregne og tolke ulike typer standardskårer
- Gjennomføre og formidle resultatene fra multippel lineær regresjonsanalyse
- Gjennomføre og formidle resultatene fra medierings- og modereringsanalyser

### Kjernebegreper

- **Aritmetisk gjennomsnitt:** Vanlig gjennomsnitt. Sum delt på antall.
- **Sann verdi (populasjonsparameter):** En populasjonsparameter er en verdi som beskriver en egenskap ved hele populasjonen, som populasjonens gjennomsnitt,

median, varians, eller proporsjon. Den sanne verdien av denne parameteren er det nøyaktige, men ofte ukjente, tallet som beskriver denne egenskapen for hele populasjonen.

I kvantitativ forskning vil man ofte forsøke å estimere den sanne verdien basert på ens utvalg.

- **Konfidensintervall:** Et intervallestimat som brukes for å indikere usikkerheten rundt et estimat av en populasjonsparameter. Det gir et (tall)område som *antas* å inneholde den sanne verdien av parameteren - med en viss konfidensnivå.

Men: Konfidensnivået (vanligvis 95%) er *ikke* sannsynligheten for at det spesifikke konfidensintervallet inneholder den sanne verdien. Dette er en vanlig misforståelse. I stedet er det en indikator på hvor ofte konfidensintervallet, konstruert over mange uavhengige prøveutvalg fra samme populasjon, vil fange den sanne verdien.

Selv om det ikke gir en sannsynlighet for en enkelt beregning, gir konfidensintervallet en nyttig indikasjon på usikkerheten rundt estimatet basert på de tilgjengelige dataene og den anvendte metoden.

- **Bootstrapping:** En statistisk metode der nye datasett simuleres med utgangspunkt i ett enkelt opprinnelig datasett. En form for resampling. Formålet er som regel å oppnå en estimering av usikkerheten rundt statistiske estimater (som f.eks. gjennomsnitt), altså å si noe om hvor sannsynlig det er at verdien du har funnet for utvalget ditt er i nærheten av den sanne verdien (populasjonsverdien).

Eksempel:

Du har et utvalg på 100 personer fra en populasjon. Du har målt en eller annen egenskap (egenskap X) ved disse 100 personene som *ikke* er normalfordelt. Du har så funnet gjennomsnittet av egenskap X i utvalget ditt og ønsker å finne ut noe om hvor sannsynlig det er at dette gjennomsnittet er representativt for populasjonen som helhet. Problemet er at de fleste tradisjonelle statistiske metoder for å beregne usikkerhet (som standardfeil og konfidensintervaller) antar normalfordeling. Derfor bruker du bootstrapping.

Hvis du skulle gjort det manuelt ville du gjort det sånn her:

1. Skrevet ned alle 100 observasjonene på lapper og lagt dem i en skål.
2. Trukket fra skålen, notert ned observasjonen, lagt lappen tilbake, trukket igjen, osv (tilfeldig utvalg med tilbakelegging/random sampling with replacement).
3. Etter å ha gjort dette 100 ganger ville du hatt ditt første simulerte utvalg.
4. Så ville du regnet ut gjennomsnittet til dette utvalget.
5. Så ville du repetert prosessen 1000-vis av ganger.

Det gidder du ikke, så du bruker PC.

PCen finner på no time gjennomsnittene til alle de 1000-vis av simulerte utvalgene. Gitt at det opprinnelige utvalget ditt var nogenlunde representativt (at fordelingen av egenskap X i utvalget ditt var ca. likt som i populasjonen som helhet) kan du nå bruke denne dataen til å regne ut for eksempel konfidensintervallet.

Bootstrapping kan også brukes til bl.a.: Hypotesetesting, modellvalidering, konstruksjonsjon av prediktive intervaller, feilkorreksjon/biasreduksjon

- **Kausalforhold:** Et forhold der en hendelse, handling, tilstand eller variabel (årsak) direkte påvirker en annen (virkning). Med andre ord, endringer i årsaken fører til endringer i virkningen. Kausalitet er grunnlaget for å si at noe "forårsaker" noe annet, og det ligger til grunn for mye av vitenskapelig forskning og teori, hvor målet er å identifisere og forstå disse forholdene.

For at et forhold skal betraktes som kausalt, må det ofte oppfylle visse kriterier, som temporalitet (årsaken må forekomme før virkningen), sammenheng (det må være en konsekvent observasjon av forholdet), og ikke-tilfeldighet (forholdet er ikke på grunn av tilfeldige, eksterne faktorer).

Å etablere kausalitet krever ofte kontrollerte eksperimenter hvor forskere manipulerer en antatt årsak for å observere effekten på en antatt virkning. I observasjonsstudier er det mer utfordrende å etablere kausale forhold på grunn av potensielle konfunderende variabler.

Kausalforhold gir grunnlaget for kausalforklaringer. Når et kausalforhold er identifisert eller antatt, søker kausalforklaringer å forklare hvorfor dette forholdet eksisterer og hvordan det fungerer.

- **Kausalforklaringer:** Kausalforklaringer er forsøk på å forklare hvorfor eller hvordan et bestemt forhold er kausalt. De gir en teoretisk eller konseptuell forståelse av mekanismene og prosessene som ligger bak kausalforholdet.
- **Variabler (kvantitative/kvalitative og uavhengige/avhengige):** I forskning og statistikk er en variabel noe som kan ta på seg forskjellige verdier. "Antall pølser solgt" kan være en variabel, det samme kan "kjønn".

"Antall pølser solgt" er en kvantitativ variabel. Disse uttrykker numeriske verdier og kan måles i tall. De kan videre deles inn i:

- Kontinuerlige variabler: Kan ta på seg ethvert verdi innenfor et gitt område - f.eks. høyde, vekt, temperatur. Du kan veie 6,9kg, 69kg, 69,69kg, 69,420kg, osv. i det

uendelige. Men du kan ikke veie -69kg. Derav "innenfor et gitt område".  
Uendelig mengde desimaler, uendelig mange gyldige verdier.

- Diskrete variabler: Tar på seg spesifikke, separate verdier - f.eks. antall pølser solgt, antall barn i en familie. Det kan ikke være 4,2 barn i en familie. Kun noen verdier gyldige.

"Kjønn" er en kvalitativ variabel. Disse beskriver kategorier eller grupper og uttrykker ikke numeriske verdier. De kan videre deles inn i:

- Nominalskala: Variabler uten naturlig rekkefølge eller rangering (f.eks. kjønn, nasjonalitet, merkenavn).
- Ordinalskala: Variabler med en naturlig rekkefølge, men uten en fast avstand mellom kategoriene (f.eks. utdanningsnivå - VGS er en høyere utdanning enn grunnskolen, men forskjellen mellom de ulike nivåene er ikke "fast").

En uavhengig variabel, ofte kalt forklarende- eller prediktorvariabel, er en variabel som forskere manipulerer eller måler for å se effekten på en avhengig variabel. Den antas å være årsaken i en årsak-virkning forhold.

I eksperimentelle design er den uavhengige variabelen den variabelen forskeren endrer for å se hva effekten blir. I observasjonsstudier tjener den som en prediktor som forskere tror kan påvirke utfallet av interesse.

Eksempler:

- I et eksperiment for å teste effekten av lys på planters vekst, er mengden lys den uavhengige variabelen.
- I en studie som ser på sammenhengen mellom trening og hjertehelse, er antall timer brukt på trening per uke den uavhengige variabelen.

En avhengig variabel, ofte referert til som utfallet eller responsvariabelen, er variabelen som forskere observerer og måler for å se endringene som kan oppstå som respons på manipulasjonen eller variasjonen i den uavhengige variabelen. Den antas å være effekten i et årsak-virkning forhold - å respondere på endringer i den uavhengige variabelen.

I stedet for å bli manipulert direkte, blir denne variabelen observert og målt.

Eksempler:

- I planteeksperimentet vil plantens veksthastighet være den avhengige variabelen som forskere måler etter å ha eksponert plantene for forskjellige lysnivåer.
- I studien om trening og hjertehelse, kan den avhengige variabelen være en måling av hjertehelse, som blodtrykk eller kolesterolnivå.

- **Konstantledd:** I statistikk og regresjonsanalyse et tall som representerer verdien til den avhengige variabelen når alle uavhengige variabler er lik null. Konstantleddet er også kjent som interceptet eller skjæringspunktet.
- **Moderatorvariabel:** En moderatorvariabel er en kvalitativ (f.eks., kjønn, rase) eller kvantitativ (f.eks., alder, inntekt) variabel som påvirker styrken eller retningen av forholdet mellom en uavhengig og en avhengig variabel.

Eksempel: I en studie av effekten av trening på mental helse, kan kjønn (W) fungere som en moderatorvariabel. Kanskje effekten av trening på mental helse er sterkere for menn enn for kvinner, eller omvendt. I dette tilfellet endrer kjønn styrken av forholdet mellom trening og mental helse.

Moderatorvariabler anngår de eksterne forholdene eller gruppene som endrer hvordan X og Y er relatert.

- **Moderatoreffekt (interaksjonseffekt):** Den effekten som oppstår når det en moderatorvariabel inne i bildet. Også kjent som en interaksjonseffekt i regresjonsanalyse.

Eksempel: Du studerer hvordan opplevd grad av støtte på jobben og stress påvirker jobbprestasjon. Du finner at med lav opplevd støtte fører økt stress til lavere jobbprestasjon, mens med høy opplevd støtte fører økt stress til høyere jobbprestasjon. Du har da funnet ut at det foreligger en moderatoreffekt, hvor opplevd støtte er en moderatorvariabel som påvirker retningen av forholdet mellom stress og jobbprestasjon.

- **Mediatorvariabel:** En mediatorvariabel (M) formidler effekten av den uavhengige variabelen (X) på den avhengige variabelen (Y). Med andre ord, X fører til M, som igjen fører til Y. Denne effekten kan være hel (at all X sin effekt på Y går via M) eller delvis. Se "Mediatoreffekt" under. Mediatorvariabelen hjelper med å forstå "hvordan" eller "hvorfor" det er et forhold mellom X og Y.

Eksempel: Du studerer effekten av trening (X) på mental helse (Y). Hvis du finner ut at trening forbedrer søvnkvalitet (M), som igjen forbedrer mental helse, så fungerer søvnkvalitet som en mediatorvariabel i forholdet mellom trening og mental helse.

Mediatorvariabler anngår den interne mekanismen eller prosessen som forklarer forholdet mellom X og Y.

- **Mediatoreffekt (medieringeffekt):** en situasjon der forholdet mellom en uavhengig variabel (X) og en avhengig variabel (Y) delvis eller helt går gjennom en tredje variabel, kjent som en mediator (M). Med andre ord, mediatoreffekten oppstår når endringer i X fører til endringer i M, som igjen fører til endringer i Y.

Fullstendig mediering er når hele forholdet mellom X og Y forklares gjennom mediatoren M, og direkte effekt av X på Y blir ikke-signifikant når M er inkludert i modellen.

Delvis mediering er når mediatoren forklarer en del av forholdet mellom X og Y, men det er fortsatt en signifikant direkte effekt av X på Y.

Viktig: Å finne støtte for en mediatoreffekt betyr ikke automatisk at du har bevist et kausalforhold mellom variablene i analysen.

Å etablere kausalitet betyr å vise at en endring i en variabel (X) *fører til* (og ikke bare samvarierer med) en endring i en annen variabel (Y) - i dette tilfellet gjennom M.

Når du observerer en mediatoreffekt, observerer du samvariasjon mellom X og M, og mellom M og Y. Samvariasjon betyr bare at to variabler beveger seg sammen på en måte som kan forutses statistisk.

Selv om du finner en mediatoreffekt, kan andre uobserverte variabler påvirke både X, M, og Y. Disse tredje variablene kan skape en illusjon av kausalitet der ingen eksisterer.

Obs: Andelen forklart varians mellom to faktorer er ikke en indikasjon på om en mediatoreffekt er tilstede eller ikke.

- **Indirekte effekt:** I konteksten statistikk og forskningsmetodikk, spesielt innen medieringsanalyse, refererer indirekte effekt til effekten en uavhengig variabel har på en avhengig variabel gjennom en eller flere mediatorvariabler. Den indirekte effekten fanger opp hvor mye av den totale effekten som går gjennom mediatorvariabelen(e), i stedet for direkte fra den uavhengige variabelen til den avhengige variabelen.
- **Mediatorhypotese:** En hypotese om at hele eller deler av den uavhengige variabelen X sin effekt på den avhengige variabelen Y går gjennom en mediatorvariabel M.

Å teste denne hypotesen innebærer normalt tre trinn:

Trinn 1: Etablere at den uavhengige variabelen påvirker den avhengige variabelen.

Trinn 2: Etablere at den uavhengige variabelen påvirker mediatoren.

Trinn 3: Etablere at mediatoren påvirker den avhengige variabelen mens du kontrollerer for den uavhengige variabelen.

I tillegg vil du ofte også ønske å estimere størrelsen på mediatoreffekten. Dette kan gjøres ved å beregne den indirekte effekten av X på Y gjennom M. Moderne metoder som bruker bootstrapping tillater en robust estimasjon av denne effekten og de tilhørende konfidensintervallene.

Mediatorhypoteser brukes for å utforske og forstå de underliggende mekanismene og prosessene som ligger bak observerte forhold mellom variabler. De hjelper forskere med å bygge teoretiske modeller som forklarer hvordan forskjellige faktorer er relatert og interagerer med hverandre. I praktisk anvendt forskning kan forståelse av mediatorer informere utvikling av mer målrettede og effektive intervensjoner ved å identifisere nøkkelpunkter for inngrep.

Selv om en mediatorhypotese kan støttes av data, er slik støtte ikke i seg selv bevis for kausalitet. Ytterligere bevis, spesielt fra eksperimentelle eller longitudinelle studier, er som oftest nødvendig.

Det finnes flere statistiske metoder for å teste mediatorhypoteser, inkludert Baron og Kenny's firetrinnsprosedyre og mer moderne tilnærminger som bruker bootstrapping for å estimere indirekte effekter.

- **Mediatormodell:** Statistiske tilnærminger som brukes for å teste og forklare mediatorhypoteser/medieringsforhold (når en mediatorvariabel formidler effekten av en uavhengig variabel på en avhengig variabel).

Selv om mediatormodeller kan indikere potensielle mekanismer for en effekt, beviser de ikke i seg selv kausalitet. I mange tilfeller kan det være mer enn én mediator, og noen variabler kan fungere både som mediatorer og moderatorer. Komplekse modeller kan være nødvendige for å fullstendig forstå disse dynamikkene.

- **Baron & Kenny's firetrinnsmodell:** En klassisk mediatormodell, spesielt mye brukt innen psykologi og sosialvitenskap. Foreslått av Baron og Kenny i 1986. Skisserer de stegene en forsker bør ta for å bestemme om og i hvilken grad mediering finner sted.

Trinn 1: Etablere at den uavhengige variabelen påvirker den avhengige variabelen.

- Mål: Vis at det er en signifikant effekt fra den uavhengige variabelen (X) til den avhengige variabelen (Y).
- Hvordan: Utfør en regresjonsanalyse der Y regresjeres på X.

Trinn 2: Etablere at den uavhengige variabelen påvirker mediatoren.

- Mål: Vis at den uavhengige variabelen (X) signifikant påvirker mediatoren (M).
- Hvordan: Utfør en regresjonsanalyse der M regresjeres på X.

Trinn 3: Etablere at mediatoren påvirker den avhengige variabelen.

- Mål: Vis at mediatoren (M) signifikant påvirker den avhengige variabelen (Y), med den uavhengige variabelen (X) også inkludert i modellen.
- Hvordan: Utfør en regresjonsanalyse der Y regresjeres på både X og M.

Trinn 4: Etablere om mediatoren fullstendig eller delvis medierer forholdet mellom den uavhengige og avhengige variabelen.

- Fullstendig mediering: Hvis effekten av X på Y blir ikke-signifikant når M er kontrollert for, antyder det fullstendig mediering.
- Delvis mediering: Hvis effekten av X på Y fortsatt er signifikant, men reduseres når M er kontrollert for, antyder det delvis mediering.

En ulempe med Baron og Kenny's fire trinn er at den indirekte effekten (mediatoreffekten) må regnes ut for hånd og at den ikke signifikant testes. Dette kan gjøres med Sobel's test, men en bedre egnet metode er å benytte bootstrapping av den indirekte effekten og beregne konfidensintervall for den indirekte effekten basert på dette.

I våre dager har nyere statistiske metoder i stor grad erstattet Baron og Kenny's.

For et grundig forklart konkret eksempel, se løsningsforslag for eksamen (høst 2023, oppgave 2d).

- **Standardfeil, SEM ("standard error of measurement"):** SEM er et mål som indikerer hvor store feilmarginer vi opererer med mht. testskårer. SEM er basert på klassisk testteori som hviler på forutsetningen om at en persons observerte skåre ( $x$ ) er et produkt av «sann skåre» ( $T$ ) og tilfeldig målefeil ( $e$ ):  $X=T+e$ . SEM estimeres på grunnlag av en kombinasjon av (1) reliabiliteten til skårene på en test og (2) standardavviket til skårene på testen. Jo større standardavvik, jo større blir SEM. Men samtidig har reliabiliteten til testskåtene også mye å si, jo lavere reliabiliteten er, jo større usikkerhet/mer målefeil og dermed større SEM. Motsatt gir høy reliabilitet lavere SEM – en perfekt reliabilitet på 1 ville gitt en SEM på null.
- **Andel forklart varians (styrke):** Et mål i statistikk som beskriver hvor stor del av totalvariansen i en datasett som kan tilskrives eller forklares av en eller flere faktorer, variabler eller modeller. Dette begrepet brukes ofte innen regresjonsanalyse, faktoranalyse, og andre statistiske modelleringsmetoder for å kvantifisere effektiviteten av en modell i å forklare observerte data.

"Styrke" i en statistisk sammenheng kan referere til styrken av et forhold, effekt, eller sammenheng. Det kan handle om hvor tydelig eller signifikant en effekt eller sammenheng er.

Andelen forklart varians kan betraktes som en indikator på styrken av modellens evne til å forklare variabiliteten i den avhengige variabelen. En høyere andel forklart varians indikerer en sterkere, mer betydningsfull modell i forhold til å fange og forklare variasjonen i dataene.



- **Sumskåre (råskåre):** En total skåre man får ved å legge sammen skårene for flere individuelle elementer eller variabler (testledd). Hvis en undersøkelse for eksempel består av 10 spørsmål av typen “Hvor enig er du i følgende påstand” som skal besvares med en skåre fra 1 (“Helt uenig”) til 10 (“Helt enig”), vil en person som svarer 10 på alle spørsmålene få en sumskåre på 100.

Dette gir en “vilkårlig og ubearbeidet” skåre (sitert sensorveiledning). Derav “rå”.

Man kan regne ut en sumskåre for en hel test/undersøkelse, eller for en hvilken som helst bestemt del av en test/undersøkelse. Hvis en spørreundersøkelse for eksempel består av tre deler, som handler om respondentens holdninger til A, B og C, kan man regne ut en sumskåre for hver av de tre delene.

Dette er en vanlig måte å måle samlede holdninger, ferdigheter, egenskaper eller andre psykometriske egenskaper på. Sumskårer brukes ofte for å gi en enkel og sammenfattende indikasjon på en større mengde data, for eksempel for å sammenligne resultater over tid eller mellom grupper.

- **Standardisering:** Handler om å omforme råskårer fra et datasett til standardskårer - skårer på en standardisert skala med et kjent gjennomsnitt og standardavvik (som 0 og 1 for Z-skårer eller 50 og 10 for T-skårer).

Hvis vi bruker Z-scoring (den klart vanligste formen for standardskåre) som eksempel, så vil en omformet råskåre som lå akkurat på gjennomsnittet i datasettet ditt alltid få en Z-skåre på 0 etter standardisering (uavhengig av opprinnelig verdi), mens en score som lå ett standardavvik over snittet vil få en Z-skåre på 1 (uavhengig av opprinnelig verdi).

Dette gjøres ofte for å gjøre data fra forskjellige kilder eller målinger sammenlignbare ved å sette dem på en felles skala. På denne måten kan du for eksempel vurdere en deltagers prestasjon på en prøve som vurderer leseferdigheter, og en annen prøve som vurderer regneferdigheter. Å simpelthen sammenligne råskårene hadde ikke sagt deg stort (la oss si at personen for eksempel skåret 69 på leseferdighetstesten, og 420 på regneferdighetstesten - er det bra eller dårlig?), men ved å omforme til standardiserte skårer kunne du f.eks. sett at personen var ett standardavvik over snittet for begge prøvene (dersom 69 og 420 var ett standardavvik over snittskårene i datasettet på de to ulike prøvene).

Standardisering brukes ofte synonymt med å omforme råscoringer til Z-skårer, siden Z-skårer er den klart vanligste formen for standardskåre og ofte brukes synonymt med ordet standardskåre.

- **Standardskåre(r) (Z-skåre(r), T-skåre(r)):** En statistisk teknikk som brukes for å beskrive en skåres posisjon i forhold til gjennomsnittet i datasettet. Standardskårer er tall basert på råskårer fra en test eller måling, som har blitt konvertert til en standardisert

skala hvor gjennomsnittet og standardavviket er konstant og kjent. Dette gjør det mulig å sammenligne skårer på tvers av forskjellige individer, tester eller test-deler.

Det er flere typer standardskårer, men den klart vanligste er Z-skåre. Z-skåre og standardskåre brukes ofte synonymt.

Z-skårer har gjennomsnitt 0 og et standardavvik på 1. Det betyr at en skåre som var på snittet i datasettet du beregner fra alltid vil omgjøres til en Z-skåre på 0, mens en skåre som var ett standardavvik over snittet i datasettet du beregner fra alltid vil få en Z-skåre på 1 - uavhengig av opprinnelig råskåre.

Z-skåre beregnes ved å subtrahere gjennomsnittsskåren i datasettet ( $\mu$ ) fra enkeltskåren du vil omgjøre til en Z-skåre ( $X$ ) og dele dette på standardavviket ( $\sigma$ ) i datasettet:

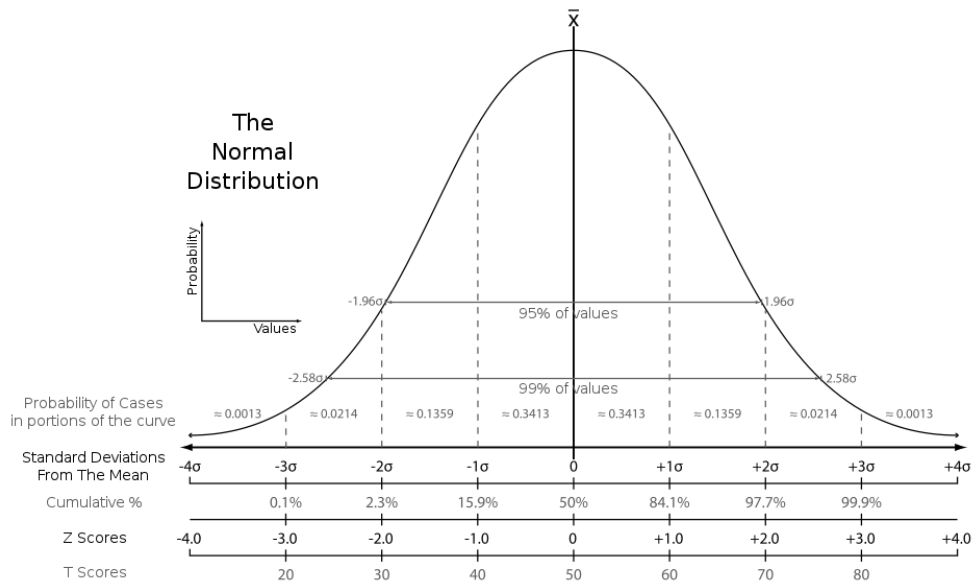
$$Z = \frac{(X - \mu)}{\sigma}$$

T-skårer ligner på Z-skårer, men omformes vanligvis til en skala med et gjennomsnitt på 50 og et standardavvik på 10. Dette gjør tallene større og ofte enklere å jobbe med. Formelen for å omdanne råskårer til en vanlig T-skåre vil være:

$$T = 50 + 10 \times \frac{X - \mu}{\sigma}$$

En annen måte å skrive dette på er  $T = 50 + (10 * Z)$ .

Dersom dataen vi har standardisert (omformet til standardskåre) var normalfordelt rundt gjennomsnittet (noe for eksempel mange, men ikke alle, menneskelige egenskaper ofte er), kan vi ved å se på standardskåren si noe hvor mange prosent av dataen skåren er høyere eller lavere enn. En Z-skåre på 1, beregnet fra et normalfordelt datasett, forteller deg at denne enkeltskåren er høyere enn omtrent 84.13% av skårene i datasettet.



- Normert(e) skåre(r):** Normerte skårer sammenligner en individuell skåre med en normgruppe eller referansepopulasjon. De gir kontekst ved å vise hvor en persons skåre faller i forhold til denne bredere gruppen. Normskårer er basert på sammenligninger med en normgruppe og trenger ikke å være standardisert i den forstand at de har et fast gjennomsnitt og standardavvik. En percentilskåre, for eksempel, forteller deg din relative posisjon i en normgruppe (for eksempel at du er i den øverste 1% av normgruppen), men ikke nødvendigvis hvor mange standardavvik du er fra gjennomsnittet.
- Stanine-skårer:** Normert skåre. "Stanine" er en forkortelse for "standard nine" og refererer til en skåringsmetode som omformer råskårer til en 9-punkts skala, vanligvis med gjennomsnittet rundt 5. Skårer under 5 er under gjennomsnittet, mens skårer over 5 er over gjennomsnittet. Selv om skalaen i seg selv er standardisert (i den forstand at den alltid har 9 punkter), gjør det faktum at den er basert på en normgruppe det til en normert skåre. Den forteller deg hvor du står i forhold til en normgruppe eller befolkning.
- Sten-skårer:** "Sten" står for "Standard Ten" og refererer til en skåringsmetode som omformer råskårer til en 10-punkts skala. Sten-skårer brukes ofte i psykologiske tester og gir en indikasjon på hvor en person faller i forhold til en normgruppe. Som med stanine-skårer gjør normreferansen dem til normerte skårer, selv om skalaen er standardisert til 10 punkter.

Både Stanine og Sten er normerte fordi de angir en persons posisjon relativt til en normgruppe. De er standardiserte i den forstand at de bruker en fast skala, men de er ikke standardskårer i samme forstand som Z-skårer eller T-skårer, som omformer råskårer til en skala med et spesifikt gjennomsnitt og standardavvik uten nødvendigvis å referere til en normgruppe.

Så, mens de har standardiserte elementer (fast skala), er de først og fremst ansett som normerte skårer.

Pensum nevner Stanine- og Sten-skårer, men man trenger ikke gå inn på dem på eksamen. T- og Z-skårer holder, de er mest de mest brukte i psykologi.

- **Korrelasjonskoeffisienter:** Et statistisk mål som beskriver graden og retningen av et lineært forhold (korrelasjon) mellom to variabler. Det er en tallverdi som kan variere fra -1 til 1, hvor verdien indikerer styrken og retningen på sammenhengen (minus = negativ korrelasjon, når den ene øker synker den andre).
- **Reliabilitet:** Om måleverktøyet ditt (f.eks. et spørreskjema) er stabilt. Et reliabelt verktøy er fritt for støy/tilfeldige målefeil og gir dermed samme resultat hver gang ved måling av det samme. Et reliabelt verktøy for måling av høyde vil f.eks. produsere samme resultat ved gjentatt måling av samme person (gitt at ikke personen har vokst i mellomtiden).

Reliabilitet er ikke det samme som validitet - at du faktisk måler det du tror du måler. Du kan ha et verktøy som du tror måler høyde, som gir det samme resultatet hver gang ved måling av samme person, men så viser det seg at det faktisk er vekt verktøyet ditt har målt hele tiden.

Formell definisjon: Reliabilitet omhandler andelen av total varians i en (sum)skåre som kan tilskrives varians i sanne skårer.

*“Et reliabelt instrument er et som oppfører seg på en forutsigbar måte, fri for tilfeldig målefeil.”* - Forelesningslide

Det finnes ulike måter å vurdere reliabilitet på. Du må kunne følgende (se definisjoner under):

- Test-retest reliabilitet
  - Alternativ form
  - Split-half reliabilitet
  - Intern konsistens (Chronbachs alpha)
- **Reliabilitetsanalyse:** En prosess man går gjennom for komme frem til estimert reliabilitet.

I løpet av en reliabilitetsanalyse vurderes typisk:

- Metoder for reliabilitetsmåling: Man velger en egnet metode for å måle reliabiliteten til testen/verktøyet. Dette kan inkludere bruk av Cronbach's alpha, test-retest korrelasjon, inter-rater korrelasjon eller andre relevante teknikker, avhengig av formålet med testen og typen data som samles inn.
  - Datainnsamling ("recording"): Man samler inn data ved å administrere testen til deltakerne i studien eller ved å observere målinger, avhengig av hva som måles.
  - Statistisk analyse: Man bruker de valgte statistiske metodene for å beregne reliabilitetsmålet, for eksempel Cronbach's alpha.
  - Tolkning og rapportering: Man tolker resultatene av reliabilitetsanalysen og rapporterer dem i forskningsrapporter eller vitenskapelige publikasjoner for å informere andre om påliteligheten til testen eller instrumentet som ble brukt.
- **Estimert reliabilitet:** Et estimat (obs!) på hvor reliabelt ditt verktøy/test/målemetode er. Basert på statistiske analyser. Tar sikte på å kvantifisere (tallfeste) graden av konsistens og stabilitet i testresultatene. Uttrykkes vanligvis som en prosentandel eller en desimal mellom 0 og 1 (reliabilitetskoeffisient), der 1 indikerer full pålitelighet. For eksempel, en estimert reliabilitet på 0,85 betyr at 85% av variasjonen i testresultatene antas å være på grunn av faktorer som testen er ment å måle, mens 15% er på grunn av tilfeldige feil eller andre faktorer.
  - **Reliabilitetskoeffisient:** En tallfesting av estimert reliabilitet. Reliabilitetskoeffisienten vil variere fra 0 (fravær av reliabilitet) til 1 (perfekt reliabilitet).
  - **Test-retest reliabilitet:** Dette refererer til graden av konsistens mellom resultatene fra to tester på samme prøveperson eller -enhet, som utføres med en viss tidsintervall mellom hver test.

Styrker:

- Enkel å forstå og implementere.
- God for trekk som er stabile over tid.

Svakheter:

- Ikke ideell for egenskaper som kan forventes å endre seg over tid (f.eks. humør).
  - Resultatene kan påvirkes av hukommelseeffekter hvis testene er for nær hverandre i tid.
- **Alternativ form:** Denne metoden bruker to ulike, men likeverdige versjoner av en test. Korrelasjonen mellom resultater fra de to versjonene brukes for å vurdere reliabiliteten.

Styrker:

- Reduserer effekten av hukommelse og øving, spesielt relevant for tester tatt kort tid fra hverandre.

Svakheter:

- Utfordrende å designe to tester som er virkelig likeverdige.
  - Kan være ressurskrevende å utvikle og administrere to separate, men likeverdige tester.
- **Split-half reliabilitet:** Dette refererer til graden av konsistens mellom resultatene fra to halvpartar av en test, der de to halvdelene skal være like i vanskelighetsgrad og innhold. Spearman-Brown-formelen brukes ofte for å justere for at hver halvdel kun er halvparten så lang som hele testen.

Styrker:

- Enkel å utføre med bare én testadministrasjon.

Svakheter:

- Resultatet avhenger av hvordan testen er delt.
  - Kan undervurdere reliabiliteten hvis de to halvdelene ikke er perfekt likeverdige.
- **Intern konsistensbasert reliabilitet (Chronbachs alpha):** Dette refererer til graden av konsistens (korrelasjon) mellom resultatene av forskjellige spørsmål eller deler av en test som antas å måle det samme konseptet. Måler hvor godt elementene i en test (indikatorer) korrelerer med hverandre og danner en konsistent skala. Beregnes basert på gjennomsnittlig inter-item korrelasjon.

Hvis vi bruker Big 5 som eksempel, vil det være høy grad av korrelasjon mellom hvordan personer besvarer spørsmålene ment å avdekke ekstroversjon. Disse spørsmålene vil da ha en høy Chronbachs alpha (som er navnet på denne formen for reliabilitetskoeffisient).

Generelt sett anses en alfa over .7 som akseptabel, over .8 som god, og over .9 som utmerket i sosialvitenskapelig forskning, selv om disse terskelverdiene kan variere avhengig av konteksten og formålet med målingen.

Obs: En høy skåre er ikke det samme som at det er bevist at det faktisk bare er én underliggende (latent) variabel som måles.

Styrker:

- Gir en god indikasjon på hvor godt et sett med elementer måler et enkelt konstrukt.
- Nyttig når det ikke er mulig eller praktisk å administrere testen flere ganger.

Svakheter:

- Alfa øker med antall elementer, noe som kan gi en illusjon av høyere reliabilitet.
- Høy alfa indikerer ikke nødvendigvis et unidimensjonalt konstrukt (kan være høy på grunn av multidimensjonalitet).
- **Inter-rater reliabilitet:** Dette refererer til graden av konsistens mellom resultatene fra to eller flere forskjellige observatører som utfører samme vurderingsoppgave. Måler graden av enighet mellom to eller flere bedømmere/ratere. Ofte brukt når subjektiv vurdering er involvert i scoringen.

Styrker:

- Viktig for å sikre at testresultatene er konsistente på tvers av forskjellige bedømmere.
- Kan brukes til å forbedre treningsprotokoller for bedømmere.

Svakheter:

- Krever flere ratere, noe som kan være ressurskrevende.
- Reliabiliteten kan påvirkes av raternes trening og erfaring.
- **Korrelasjonsmatrise:** En korrelasjonsmatrise er en statistisk teknikk brukt for å få oversikt over forholdene mellom to og to variabler i et datasett. Matrisen er en tabell hvor hver celle inneholder en korrelasjonskoeffisient mellom variabelen på X-aksen og variabelen på Y-aksen.

I eksempelet under kan vi se at korrelasjonen mellom variabel 7 og 1 i dette datasettet er 0.642578. Det betyr at når den ene øker med 1, øker den andre med 0.642578. Korrelasjonskoeffisienten for 9 og 2 er -0.383585, noe som betyr at når den ene stiger med 1, synker den andre med 0.383585. Korrelasjonsmatriser brukes oftest til å lage regresjonsmodeller.

	0	1	2	3	4	5	6	7	8	9
0	1.000000	0.347533	0.398948	0.455743	0.072914	-0.233402	-0.731222	0.477978	-0.442621	0.015185
1	0.347533	1.000000	-0.284056	0.571003	-0.285483	0.382480	-0.362842	0.642578	0.252556	0.190047
2	0.398948	-0.284056	1.000000	-0.523649	0.152937	-0.139176	-0.092895	0.016266	-0.434016	-0.383585
3	0.455743	0.571003	-0.523649	1.000000	-0.225343	-0.227577	-0.481548	0.473286	0.279258	0.446650
4	0.072914	-0.285483	0.152937	-0.225343	1.000000	-0.104438	-0.147477	-0.523283	-0.614603	-0.189916
5	-0.233402	0.382480	-0.139176	-0.227577	-0.104438	1.000000	-0.030252	0.417640	0.205851	0.095084
6	-0.731222	-0.362842	-0.092895	-0.481548	-0.147477	-0.030252	1.000000	-0.494440	0.381407	-0.353652
7	0.477978	0.642578	0.016266	0.473286	-0.523283	0.417640	-0.494440	1.000000	0.375873	0.417863
8	-0.442621	0.252556	-0.434016	0.279258	-0.614603	0.205851	0.381407	0.375873	1.000000	0.150421
9	0.015185	0.190047	-0.383585	0.446650	-0.189916	0.095084	-0.353652	0.417863	0.150421	1.000000

- **(Målings)validitet:** Hvor godt et verktøy/test/forskningsmetode faktisk måler det det hevder å måle.

Formelt målingsvaliditet, kalles ofte bare validitet.

Et verktøy/test/forskningsmetode kan ha høy reliabilitet, men lav validitet. En badevekt som måler brukerens høyde istedenfor vekt har svært lav validitet, men så lenge den konsekvent treffer på høyden, er reliabiliteten høy.

- **Innholdsvaliditet:** Et mål på hvor godt et testinstrument, som en spørreundersøkelse, intervju eller annen test, dekker hele omfanget av det konseptet det er ment å måle. Ofte blir innholdsvaliditet vurdert gjennom en prosess der eksperter på det aktuelle området gjennomgår testen for å sikre at den dekker alle relevante deler av konstruktet. Dette kan inkludere å vurdere om noe viktig mangler og om det er elementer som ikke bør være inkludert.
- **Kriterievaliditet (prediktiv- og samtidig-):** Prediktiv og samtidig kriterievaliditet er to former for kriterievaliditet, som handler om å vurdere hvor godt et måleinstrument korrelerer med et relevant ytre kriterium. Disse formene for validitet er spesielt viktige når målet er å demonstrere at et instrument kan gi nyttige indikasjoner på bestemte utfall eller atferd.

Prediktiv kriterievaliditet

- Definisjon: Prediktiv kriterievaliditet refererer til hvor godt et måleinstrument kan forutsi utfall eller atferd på et senere tidspunkt. Det er en vurdering av instrumentets evne til å forutsi fremtidige hendelser, tilstander eller oppførsel.
- Eksempel: Hvis du har en test som måler studieferdigheter hos studenter, ville prediktiv validitet vurderes ved å korrelere testresultatene med studentenes fremtidige akademiske prestasjoner, som karakterer eller eksamensresultater,



oppnådd noen tid etter at testen ble tatt. Høy korrelasjon ville indikere høy prediktiv kriterievaliditet.

### Samtidig kriterievaliditet

- Definisjon: Samtidig kriterievaliditet (noen ganger kalt konkurrent validitet) vurderer hvor godt et måleinstrument korrelerer med et annet mål som er tatt på omtrent samme tidspunkt. Det fokuserer på instrumentets evne til å reflektere nåværende tilstander eller atferd.
- Eksempel: For en test som måler nåværende nivå av depresjon, ville samtidig kriterievaliditet innebære å korrelere resultatene fra testen med resultatene fra et annet etablert mål for depresjon tatt samtidig. En sterk korrelasjon ville indikere at det nye målet har høy samtidig kriterievaliditet.

Den primære forskjellen mellom de to ligger med andre ord i når kriteriedataene samles inn. Prediktiv validitet ser fremover i tid, mens samtidig validitet ser på nåværende korrelasjoner. Valget mellom prediktiv og samtidig kriterievaliditet avhenger av om måleinstrumentets formål. Hvis formålet er å forutsi fremtidige utfall, er prediktiv validitet mer relevant. Hvis målet er å vurdere nåværende tilstander eller forhold, er samtidig validitet mer hensiktsmessig.

- **Begrepsvaliditet (konstruktvaliditet):** Handler om hvor godt et måleinstrument faktisk måler det teoretiske konstruktet eller begrepet det hevder å måle.

*“Begrepsvaliditet er en form for målingsvaliditet som viser til om (skårene på) testen opptrer i tråd med vårt teoretiske grunnlag.” - Sensorveiledning*

Pensum fokuserer på to metoder for å vurdere begrepsvaliditet: konvergent validitet og diskriminant validitet.

- Konvergent validitet refererer til graden hvorpå to mål som antas å måle det samme konstruktet, er i faktisk korrelasjon med hverandre. For eksempel, hvis du har utviklet en ny test for å måle angst, bør denne testen korrelere positivt med andre etablerte angsttester. Hvis din nye angstsкала viser høy korrelasjon med en annen anerkjent angstsкала, indikerer dette god konvergent validitet.
- Diskriminant validitet er den graden hvorpå et måleinstrument ikke korrelerer med andre variabler og konstrukter som det teoretisk ikke skal ha sammenheng med. Fortsetter med angsteksemplet: Mens din angstsкала bør korrelere med andre angstsкалаer (konvergent validitet), bør den ikke korrelere med målinger av ulike konstrukter, som for eksempel fysisk koordinasjon. Hvis din angstsкала viser lav korrelasjon med målinger som er irrelevante for angst, viser det god diskriminant validitet.

En statistisk teknikk som brukes til å vurdere begrepsvaliditeten ytterligere er multitrait-multimethod (MTMM) analyse. Den vektlegges mindre av pensum. Denne metoden er basert på ideen om at en grundig vurdering av et konstrukt krever at det måles på flere måter (multimethod) og at flere konstrukt (multitrait) bør måles samtidig for å vurdere deres forhold til hverandre. I en MTMM-matrise måles flere trekk (traits) ved bruk av flere metoder (methods). For eksempel kan et trekk som angst måles via selvrapporing, observasjon, og fysiologiske målinger. Samtidig kan andre trekk som depresjon eller lykke også måles ved hjelp av disse metodene. Dermed gir denne tilnærmingen data som gjør det mulig å vurdere både konvergent og diskriminant validitet samtidig. Da MTMM er komplekst og ressurskrevende, blir den ofte mindre brukt enn konvergent og diskriminant validitet.

På eksamen er det ønskelig, men ikke påkrevd, å trekke inn faktoranalyse i en definisjon av begrepsvaliditet - da særlig vektlegge at faktoranalyse kan hjelpe en med å forstå den interne strukturen til måleinstrumentet ens og å bestemme om testen måler én eller flere latente variabler (faktorer).

- **Måleinstrumentets interne struktur:** refererer til hvordan forskjellige elementer eller spørsmål (ofte kalt "items") innenfor et måleinstrument er organisert og korrelerer med hverandre for å måle et eller flere underliggende konstrukt eller dimensjoner. Å forstå et måleinstrument sin interne struktur er viktig for å sikre at det måler det det er ment å måle (konstruktvaliditet) og at det gjør det på en konsistent og pålitelig måte.

Når man utforsker måleinstrumentets interne struktur, ser man på flere aspekter:

1. Dimensjonalitet: Forskere ønsker å vite om måleinstrumentet måler ett enkelt konstrukt (unidimensjonalt) eller flere relaterte konstrukt (multidimensjonalt). For eksempel, en personlighetstest kan ha flere dimensjoner som hver representerer forskjellige trekk som ekstraversjon, omgjengelighet, og åpenhet.
2. Item-korrelasjoner: Dette handler om hvordan individuelle items korrelerer med hverandre. Forventningen er at items som er ment å måle det samme konstruktet skal ha moderat til høy positiv korrelasjon. Hvis visse items ikke korrelerer som forventet, kan det være tegn på at de ikke passer godt inn i det overordnede konstruktet.
3. Faktorstruktur: En vanlig metode for å utforske måleinstrumentets interne struktur er faktoranalyse (inkludert eksplorerende og bekreftende faktoranalyse). Faktoranalyse hjelper med å identifisere underliggende faktorer (eller dimensjoner) som items grupperer seg rundt, noe som gir innsikt i måleinstrumentets struktur og dimensjonalitet.
4. Item-respons-teori (IRT) analyser: IRT er en annen tilnærming som ser på forholdet mellom testtakernes trekknivå og sannsynligheten for en bestemt respons på hvert item. IRT-modeller kan gi detaljert informasjon om items' egenskaper, som deres vanskelighetsgrad og diskriminerende kraft.

Å forstå måleinstrumentets interne struktur er avgjørende for flere grunner:

- Forbedre validitet: Det sikrer at alle items bidrar til en konsistent måling av det tiltenkte konstruktet.
- Reliabilitetsvurdering: Det hjelper med å vurdere hvor pålitelig instrumentet måler konstruktet (for eksempel gjennom Cronbachs alfa for interne konsistens reliabilitet).
- Teoretisk forståelse: Det gir innsikt i naturen og strukturen av konstruktet som måles, og kan bidra til videre teoretisk utvikling.

Å nøye vurdere og forstå et måleinstrument sin interne struktur er derfor en sentral del av psykometrisk evaluering og testutvikling.

- **Subskala:** En mindre, distinkt del av en større skala eller måleinstrument. En gruppe av relaterte elementer (spørsmål, påstander, indikatorer) innenfor en større skala som sammen måler en spesifikk, underliggende dimensjon eller faktor av det overordnede konstruktet. Tenk Big 5 - hvert trekk har underliggende subskalaer.
- **Faktoranalyse:** Faktoranalyse er en statistisk metode som brukes til å undersøke underliggende strukturer i et datasett, vanligvis for å identifisere grupper av relaterte variabler, kjent som faktorer, som bidrar til mønstre i dataene.

Eksempel: Big 5. Der brukte man faktoranalyse for å komme frem til de 5 dimensjonene, f.eks. ved at man fant korrelasjoner av typen "folk som beskriver seg selv som "pessimistiske", beskriver seg konsekvent *ikke* som "lettsindige"". Disse korrelasjonene, i stor skala, viste at både "pessimistisk" og "lettsindig" er adjektiv som beskriver det underliggende konstruktet "Nevrotisisme".

Faktoranalyse kan brukes til å vurdere begrepsvaliditeten til måleinstrumenter, særlig når det gjelder å forstå den interne strukturen og å bestemme om testen måler én eller flere latente variabler (faktorer).

- Identifisere latente faktorer: Faktoranalyse hjelper med å utforske den underliggende dimensjonaliteten av et sett med målinger. Ved å analysere mønstre av korrelasjoner mellom målevariablene, kan forskere identifisere grupper eller "faktorer" som ser ut til å reflektere underliggende konsepter eller trekk.
- Undersøke interne strukturer: Ved å vurdere hvordan individuelle variabler laster på de ulike faktorene, kan forskere avgjøre om måleinstrumentet er konsistent internt. For eksempel, hvis et måleinstrument er designet for å måle et enkelt konstrukt (som angst), forventer man at alle eller de fleste av elementene i instrumentet skal laste betydelig på én enkelt faktor.
- Teoritestende: Faktoranalyse kan også være teoritestende. Hvis det finnes en hypotese om at et måleinstrument skal reflektere flere underliggende konstrukt

(som ulike typer angst), kan forskere bruke konfirmatorisk faktoranalyse (CFA) til å teste hvor godt de observerte dataene passer med denne hypotesen.

- Forbedring, modifisering og validering: Basert på resultatene fra faktoranalysen, kan forskere foreta modifikasjoner av måleinstrumentet. Dette kan inkludere å fjerne elementer som ikke laster godt på noen faktor, eller de som laster på mer enn én faktor, noe som indikerer at de kanskje måler mer enn ett konstrukt.

La oss for eksempel si at du har en test designet for å måle stress. Etter å ha utført en faktoranalyse, finner du at spørsmålene grupperer seg i tre faktorer: arbeidsrelatert stress, sosialt stress, og stress på grunn av personlige forhold. Dette indikerer at testen kanskje måler tre separate, men relaterte, latente variabler snarere enn ett enkelt stresskonstrukt. Dette kan lede til en revisjon av testen eller til en ny forståelse og definisjon av konstruktet som måles.

Ved å bruke faktoranalyse kan forskere dermed få innsikt i måleinstrumentets struktur og dets evne til å måle ønskede konstrukt, noe som er avgjørende for å sikre høy begrepsvaliditet.

- **Faktorladning:** En numerisk verdi som representerer styrken og retningen av forholdet mellom en observasjon (for eksempel en spørreundersøkelsesvare) og en faktor (en underliggende dimensjon). Det er et mål på hvor mye en variabel "laster" eller bidrar til en faktor. De indikerer hvor mye av variansen i hver variabel som kan forklares av hver faktor og bidrar til å definere betydningen og sammensetningen av de latente faktorene.

Lik korrelasjonskoeffisienter kan faktorladninger variere fra -1 til 1. En ladning nær 1 eller -1 indikerer at variabelen er sterkt assosiert med faktoren, mens en ladning nær 0 indikerer liten eller ingen assosiasjon. Positive ladninger indikerer at variabelen øker med faktoren, mens negative ladninger indikerer at variabelen avtar når faktoren øker.

I en faktoranalysemodell representeres ladningene i en ladningsmatrise, hvor hver rad representerer en variabel, og hver kolonne representerer en faktor. Ladningene brukes til å tolke faktorene. For eksempel, hvis flere variabler som måler tilfredshet har høye ladninger på en bestemt faktor, kan du tolke denne faktoren som en "tilfredshetsfaktor".

Typer faktorladninger:

- Utforskende faktoranalyse (EFA): I EFA er målet å oppdage strukturen og antall faktorer som best forklarer koblingene mellom variablene. Faktorladningene her hjelper med å identifisere og tolke disse faktorene.
- Konfirmatorisk faktoranalyse (CFA): I CFA brukes faktorladninger til å teste teoretisk forventede forhold mellom observerte variabler og deres underliggende faktorer.

- **Kommunalitet:** I konteksten av faktoranalyse et mål på den variansen i en observasjon (for eksempel et spørsmål i en spørreundersøkelse eller en test) som deles eller forklares av de felles faktorene i modellen.

I begynnelsen av en faktoranalyse, er kommunalitetene ofte antatt å være lik 1 (som antyder at all varians er felles). Etter hvert som analysen fortsetter, blir kommunalitetene justert for å reflektere hvor mye av variansen som faktisk deles med de andre variablene gjennom de felles faktorene.

Kommunalitet for en variabel er ofte representert ved  $h^2$  i faktoranalyse.

Kommunalitet gir innsikt i hvor godt en variabel passer inn i faktorstrukturen som er etablert av datasettet. En høy kommunalitet indikerer at en betydelig del av variabelens varians kan forklares av faktorene i modellen.

Selv om en variabel har høy kommunalitet, kan det fortsatt være en betydelig del av variansen som ikke forklares av modellen (unike eller feilvariens).

- **Regresjonsmodeller:** Statistiske teknikker brukt for å forstå og forutsi forholdet mellom avhengige (mål) og uavhengige (forklarer) variabler. En regresjonsmodell produserer en matematisk formel som beskriver forholdet mellom den avhengige variabelen (den du ønsker å forstå eller forutsi) og en eller flere uavhengige variabler (de faktorene du antar påvirker den avhengige variabelen). Denne formelen kan deretter brukes til å beregne eller forutsi en sannsynlig verdi for den avhengige variabelen gitt spesifikke verdier av de uavhengige variablene.
- **Regresjonsanalyse:** Å bruke data for å bygge og validere en regresjonsmodell. Mens modellen er det teoretiske verktøyet, er analysen prosessen med å anvende og vurdere dette verktøyet i praksis. Eksempel: Hvis du har data om huspriser og deres egenskaper (størrelse, beliggenhet, alder, osv.), kan en regresjonsmodell teoretisk beskrive hvordan hver egenskap forventes å påvirke prisen. Regresjonsanalysen vil da være prosessen med å faktisk beregne denne modellen med dine data, teste dens nøyaktighet, og bruke den til å forutsi prisen på hus basert på deres egenskaper.

Sentrale begreper/konsepter:

- Retning i sammenheng: Om forholdet mellom den avhengige og uavhengige variabelen er positivt eller negativt. I en positiv sammenheng øker verdien av den avhengige variabelen når verdien av den uavhengige variabelen øker. I en negativ sammenheng er det motsatt. Retningen indikeres ofte av tegnet på koeffisientene i regresjonsmodellen. Forståelse av retningen hjelper med å tolke hvordan variabler påvirker hverandre, noe som er sentralt i både modellutvikling og beslutningstaking.

- Konfundert sammenheng: Oppstår når effekten av en uavhengig variabel på den avhengige variabelen er blandet eller forvirret med effekten av en annen uavhengig variabel. Dette kan føre til at forholdet mellom den opprinnelige uavhengige variabelen og den avhengige variabelen blir feiltolket. Å identifisere og kontrollere for konfunderte variabler er derfor viktig.
- Multippel regresjonsanalyse: En utvidelse av enkel lineær regresjon som tillater inkludering av to eller flere uavhengige variabler (prediktorer) for å forutsi den avhengige variabelen. Multippel regresjonsanalyse lar en forstå hvordan flere variabler samlet påvirker den avhengige variabelen - og hvordan hver enkelt bidrar - mens man kontrollerer for effektene av de andre. Den er egnet til å analysere komplekse forhold og kan hjelpe med å identifisere og kontrollere for konfundering.
- **Regresjonskoeffisient (B):** Et tall som representerer størrelsen og retningen av forholdet mellom en uavhengig variabel og den avhengige variabelen i en regresjonsmodell.

En standardisert regresjonskoeffisient, ofte referert til som en beta-koeffisient i konteksten av lineær regresjon ( $\beta$ ), er en versjon av regresjonskoeffisienten som har blitt justert for å reflektere antall standardavvik en avhengig variabel vil endre seg som respons på en endring på ett standardavvik i den uavhengige variabelen. Denne standardiseringen tillater for sammenligning av styrken og viktigheten av hver uavhengig variabels innflytelse på den avhengige variabelen i modellen, spesielt når variablene er målt på forskjellige skalaer.

- **Statistisk signifikans:** Sannsynligheten for at en observert forskjell i dataene skyldes en reell effekt, og ikke bare tilfeldigheter. Hvis et funn er statistisk signifikant, betyr det at det er en lav sannsynlighet for at resultatet ble oppnådd tilfeldig eller ved flaks.
- **Signifikansnivå ( $\alpha$  - alfa):** Et nivå for p-verdien bestemt på forhånd som terskelen for å forkaste nullhypotesen. Hvis p-verdien er lavere enn alfa, forkastes nullhypotesen. Vanligvis brukes alfa-nivåer på 0,05 eller 0,01, dvs. 5% / 1% sjans for å gjøre type 1-feil.
- **P-verdi:** Sannsynligheten for å observere en forskjell i dataene som er like stor eller større enn det som faktisk ble observert, ved en tilfeldighet. Går fra 0 (0% sjans) til 1 (100% sjans). Hvis p-verdien er lavere enn det forhåndsbestemte signifikansnivået (alfa), forkastes nullhypotesen.
- **Signifikanstest:** Også kjent som en statistisk signifikansprøve. En statistisk prosedyre. Signifikanstester brukes til å avgjøre om forskjeller eller effekter observert i dataene sannsynligvis skyldes tilfeldig variasjon eller er "ekte" - om det finnes tilstrekkelig bevis i datamateriale til å konkludere at en bestemt betingelse eller egenskap gjelder for en

større populasjon.

Denne prosedyren er en grunnleggende del av hypoteseprøving i statistisk analyse.

Prosess:

1. Formulere Hypoteser: Bestem nullhypotesen ( $H_0$ ) og den alternative hypotesen ( $H_1$ ).
2. Velg teststatistikk: Velg en passende teststatistikk basert på dataene og det som testes (f.eks., t-test, chi-kvadrat-test, ANOVA).
3. Bestemme signifikansnivå: Velg et signifikansnivå ( $\alpha$ ), ofte satt til 0.05, som representerer terskelen for å avvise nullhypotesen.
4. Beregne P-verdien: Beregn p-verdien, som er sannsynligheten for å observere dataene, eller noe mer ekstremt, gitt at nullhypotesen er sann.
5. Konklusjon: Sammenlign p-verdien med signifikansnivået. Hvis p-verdien er mindre enn  $\alpha$ , avviser vi nullhypotesen til fordel for den alternative hypotesen.

Typer signifikanstester:

- T-test: Brukes for å sammenligne middelerverdier mellom to grupper.
  - Chi-kvadrattest: Brukes for å teste forhold mellom kategoriske variabler.
  - ANOVA: Brukes for å sammenligne middelerverdier over flere grupper.
  - Z-test: Brukes for store prøver når standardavviket er kjent.
- 
- **Kanalsammenheng**
  - **Preacher & Hayes' process-makro**