

Psykometriska grundbegrepp

PSYKOMETRISKA GRUNDBEGREPP

Psykometriska grundbegrepp med några exempel från Sensory Profile, Bayley Scales of Infant and Toddler Development, Movement Assessment Battery for Children och Clinical Evaluation of Language Fundamentals

Nedan presenteras kort några grundläggande psykometriska begrepp som förekommer i Sensory Profile, Bayley Scales of Infant and Toddler Development (Bayley), Movement Assessment Battery for Children (Movement ABC) och Clinical Evaluation of Language Fundamentals (CELF), instrument som bland annat riktar sig till fysioterapeuter och logopedier.

Råpoäng

Uppnådd total råpoäng, det vill säga en summering av poängen på testets (eller skattningsskalans) alla uppgifter, är i allmänhet inte tolkningsbar. För att resultatet ska vara möjligt att tolka måste råpoängen omvandlas till standardpoäng. En sådan omvandling görs genom en statistisk procedur som kallas skaltransformation, där fördelningen av råpoäng överförs till en standardskala. I testmanualerna finns tabeller för enkel omvandling av råpoäng till standardpoäng i de fall en papper och penna-version av testet administrerats. Om testet administrerats eller scoras digitalt sköter programmet omvandlingen. Det finns dock instrument – som exempelvis Sensory Profile – där tolkningen sker på råpoängsnivå och man använder sig av så kallade "cut-off-gränser", se nedan för en beskrivning.

Standardpoäng och standardskalor

Det finns olika typer av standardpoäng (ibland benämnda normativa poäng); några av de vanligaste är indexpoäng, T-poäng, skalpoäng, staninepoäng och stenpoäng. Först när råpoängen omvandlats till standardpoäng börjar det bli möjligt att få en uppfattning om vad poängen egentligen betyder, och hur det enskilda resultatet ska tolkas.

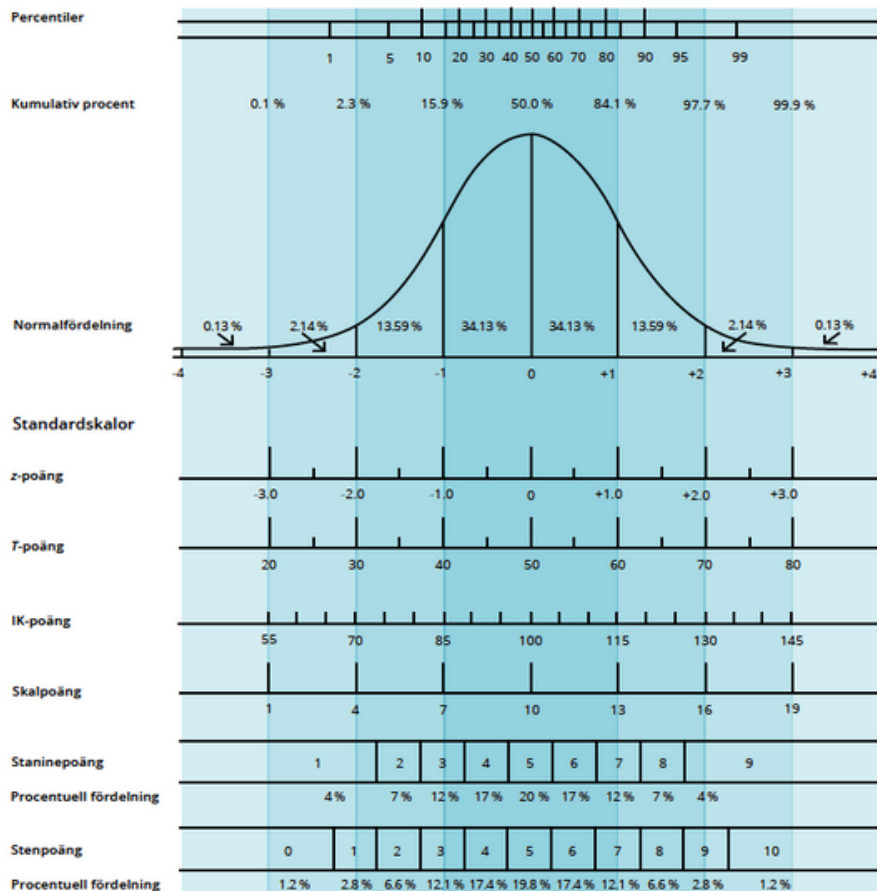
Ofta kan flera standardpoäng (exempelvis skalpoäng som ofta används på deltestnivå) i ett test summeras för att därefter omvandlas till en annan, mer sammanfattande standardpoäng. I exempelvis Bayley summeras skalpoängen på varje deltest till mer övergripande indexpoäng för de tre skalorna Kognition, Språk och Motorik. Indexpoängen används sedan för normjämförelser. Även i CELF och Movement ABC görs liknande summeringar för att ge användaren en sammanfattande, global bild av hur barnets språk och motorik utvecklats.

Standardpoäng är lätta att jämföra sinsemellan då de alla är poäng på standardskalor. Standardskalor utgår från normalfördelningen och alla standardskalor beskriver graden av "avvikelse" i det individuella resultatet i termer av avstånd från medelvärdet.

PSYKOMETRISKA GRUNDBEGREPP

Normalfördelning

Många mänskliga egenskaper är normalfördelade. Om man mäter en viss egenskap hos ett stort antal människor kommer de flesta att hamna någonstans i mitten, runt medelvärdet. Färre personer har extremt mycket, eller extremt lite, av en viss egenskap. De olika standardskalorna bygger alla på normalfördelningen. I bilden nedan blir det tydligt att merparten av alla observationer – det vill säga merparten av alla mätningar, drygt 68 %, av en viss egenskap hamnar runt medelvärdet. En del lite under, en del lite över. Ju längre ut i normalfördelningens "svansar" man kommer, desto färre observationer.



PSYKOMETRISKA GRUNDBEGREPP

Normer och normgrupper

För att kunna tolka en enskild persons resultat på ett test måste man sätta resultatet i relation till något. Hur högt eller lågt ett resultat är kan bara bestämmas i relation till hur andra har presterat. Därför har nästan alla test normer, det vill säga data insamlade från en stor mängd människor – barn eller vuxna beroende på test. När man jämför det enskilda resultatet med normerna får man en bild av hur det förhåller sig till vad andra personer presterat. I en del test finns olika normgrupper så att man kan välja den jämförelsegrupp som innehåller personer i samma ålder som den person man vill jämföra. Hur många normgrupper som finns, och hur de är indelade, varierar mellan olika test. I exempelvis Bayley finns väldigt många olika normgrupper baserade på ålder, eftersom utvecklingen av de olika förmågor testet mäter går snabbt hos små barn, medan en skala avsedd för vuxna kan ha betydligt färre normgrupper.

Standardfel och konfidensintervall

Poäng på ett test är alltid en uppskattning av en persons "sanna" poäng. De uppnådda poängen reflekterar resultatet i kombination med ett mätfel (standardfel). Sanna poäng representeras mer adekvat genom att man etablerar ett konfidensintervall runt testresultatet. Ett konfidensintervall är ett poängintervall inom vilket de sanna poängen mest sannolikt faller. Konfidensintervall är ett sätt att uttrycka precisionsgraden i de uppnådda poängen, och fungerar som en påminnelse om att alla resultat på test och skattningsskalor är behäftade med mätfel. I testmanualernas normtabeller finns i allmänhet uppgift om vilket poängintervall som utgör konfidensintervallet. Det är det intervallet, snarare än exakta poäng, som ska användas när testresultat rapporteras. Den som använder Q-global eller Q-interactive får konfidensintervallen presenterade i de resultatrapporter som genereras från systemen.

Percentiler

I såväl Bayley som i Movement ABC och CELF presenteras i tillägg till standardpoäng även percentiler. Percentiler representerar procentandelar av normgruppen som uppnådde ett högre eller lägre resultat än en given råpoäng. Percentilvärdena sträcker sig mellan 1 och 99 och kan sägas beskriva hur vanligt eller ovanligt förekommande ett visst resultat är i normgruppen. Ett enskilt testresultat som exempelvis hamnar på den 65:e percentilen betyder att resultatet är lika högt som, eller högre än, de resultat som 65 % av personerna i normgruppen fick. Eller omvänt, lägre än 35 % av jämförelsegruppen. Det kan vara lätt att förstå men man måste vara försiktig när man tolkar percentiler. Percentilerna i en normalfördelning tenderar att samlas i mitten av fördelningen. För personer som presterar genomsnittligt innebär det att en förändring på en eller två råpoäng kan ge en stor förändring uttryckt i percentilvärden. För personer som är antingen hög- eller lågpresterande medför en motsvarande förändring i råpoäng däremot inga påtagliga effekter på percentilvärdet.

PSYKOMETRISKA GRUNDBEGREPP

Diskrepansanalyser

Både Bayley och CELF innehåller många deltest. För att avgöra hur mycket resultatet på deltesten skiljer sig åt, och om den skillnaden är viktig, kan man genomföra så kallade diskrepansanalyser. I CELF kan diskrepansanalyser ge viktig information om eventuella ojämnheter i ett barns språkutveckling, och i Bayley kan användare bedöma om exempelvis utvecklingen av kognitiva förmågor och motoriska förmågor håller jämna steg med varandra eller inte. Hur diskrepansanalyser genomförs beskrivs i testmanualerna i de fall testen administreras i papper och penna-version, eller genomförs automatiskt i Q-interactive och Q-global vid digital administrering.

Åldersekvivalenter

Många test, inklusive Bayley, presenterar åldersekvivalenter. En åldersekvivalent representerar den genomsnittliga åldern, uttryckt i månader, för vilken en given total råpoäng är typisk. Åldersekvivalenter kan vara lätta att förstå och kommunicera, men de ska användas med mycket stor försiktighet. De ger ingen information om hur resultatet förhåller sig till andras i samma ålder, och de är känsliga för små förändringar i råpoäng.

Cut off-poäng

I Sensory Profile tolkas råpoängen mot uppställda så kallade "cut off-gränser", det vill säga poänggränser som markerar olika grad av samstämmighet med normgruppen. Är de uppnådda råpoängen desamma som de flesta andras, eller avviker de? Och i så fall, hur mycket eller litet? Cut off-poäng baseras ofta på ett externt kriterium och markerar övergången från en kvalitativ kategori till en annan. I Sensory Profile tjänar standardavvikelsen som detta kriterium. Standardavvikelsen (SD) är den genomsnittliga avvikelsen från medelvärdet i en serie observationsvärden. Poäng som ligger inom cut off-intervallet -1 till $+1$ SD representerar i den teoretiska normalfördelningen cirka 68 % av populationen. På motsvarande sätt innebär ett resultat där en person hamnat över en cut off-gräns på 2 SD att personen tillhör en minoritet i normalfördelningen, närmare bestämt får ungefär 2 % ett sådant högt värde. I Sensory Profile klassificeras en sådan person med etiketten "Mycket mer än de flesta andra" för det aktuella måttet.

Reliabilitet

Reliabiliteten i test påvisar med vilken precision eller tillförlitlighet testets eller skattningsskalans påståenden/frågor/uppgifter fångar det som ska mätas. Reliabilitetsåttet anges som en korrelationskoefficient som kan variera mellan 0 och 1. Det finns olika metoder att mäta reliabilitet. Den vanligaste är den som mäter sambanden mellan en skalas olika uppgifter, och som alltså ger ett mått på skalans interna konsistens.

PSYKOMETRISKA GRUNDBEGREPP

Reliabiliteten bör ligga på .70 eller högre för att det ska anses vara tillräcklig precision i mätningen. Men, i de fall reliabiliteten är mycket hög riskerar man att missa bredden i den mätta egenskapen eller förmågan; mätområdet blir för snävt. Andra vanliga reliabilitetsmått är testning inom en och samma person över tid, så kallad test-retest, eller att frågornas konsistens prövas mellan olika skattare, så kallad interbedömarreliabilitet.

Validitet

Ett tests validitet är dess viktigaste egenskap. Att ett test har validitet betyder att det mäter vad det ska mäta, och inte något annat. I motsats till reliabilitet är det svårt att sätta en siffra på validiteten. Det finns olika typer av validitet. Ofta finner man i testets dokumentation uppgifter om hur olika särskilda grupper – grupper med personer som redan konstaterats ha olika typer av svårigheter (exempelvis med sin språkliga eller motoriska utveckling) – har presterat på testet. När man utvecklar ett test samlar man data från sådana grupper för att se om testet kan skilja mellan dem och personer utan motsvarande svårigheter. Andra typer av validitet har att göra med huruvida de frågor som finns i testet är relevanta för den egenskap som ska mätas, samt om testet mäter ett någorlunda enhetligt begrepp eller delaspekter av ett begrepp. En statistisk metod benämnd faktoranalys används i allmänhet för att undersöka strukturen i testet. Syftet med en faktoranalys är att studera eventuella bakomliggande (så kallade latent) faktorer på grundval av en analys av sambanden mellan observerade (så kallade manifesta) variabler.

Normering, normvalidering och översättning

Processen med att samla in jämförelsedata till ett instrument kallas normering. Normering är ett grannlaga och krävande arbete – en stor mängd data, representativ för den population i vilken testet ska användas, måste samlas in. Hänsyn måste tas till en rad demografiska variabler (som ålder, kön och utbildningsnivå) för att instrumentets normer ska utgöra en adekvat spegling av populationen och kunna användas för jämförelseändamål. I vissa fall och för vissa instrument kan man i stället göra en så kallad normvalidering, vilket innebär att en mindre datamängd samlas in för att undersöka relevansen av den ursprungliga normuppsättningen (som ofta är amerikansk eller brittisk). Och, i vissa fall kan ett test publiceras som en kulturellt anpassad översättning av originalversionen. De görs i första hand när det som ska mätas kan antas vara mindre känsligt för språkliga och kulturella inflytanden, som exempelvis motorisk utveckling.

Niklas Hansen, fil.dr, projektledare, R&D, Pearson Clinical

Katarina Forssén, leg psykolog, projektledare, R&D, Pearson Clinical



www.pearsonclinical.se