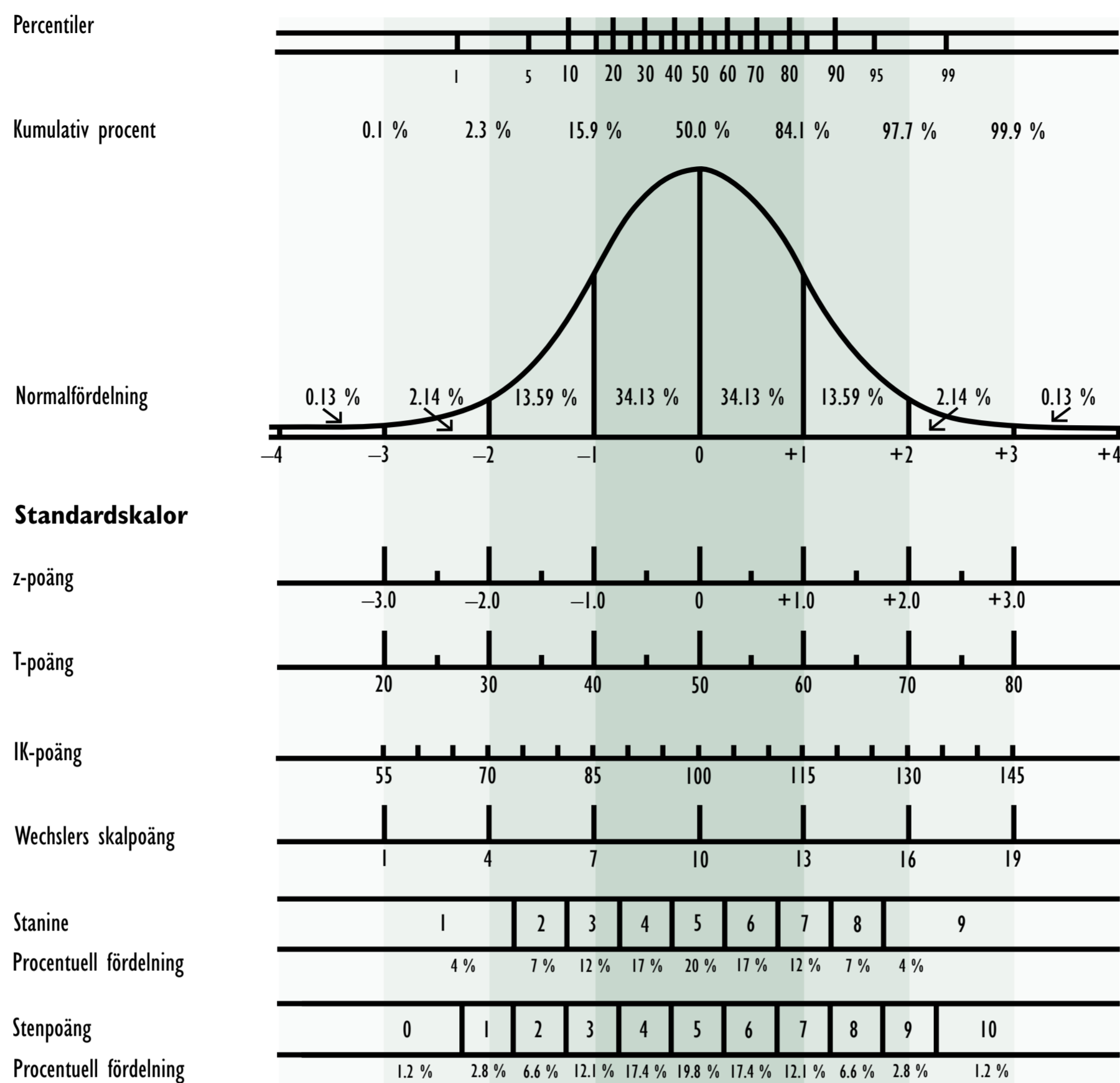


Standardskalor



Centrala begrepp i psykologisk mätning

Reliabilitet

Olika metoder kan tillämpas för att uppskatta med vilken precision en skala mäter det som avses. Vilken eller vilka metoder som är mest lämpliga är avhängigt skalans konstruktion. För att estimera reliabiliteten i en skala beräknas ett sambandsmått, en korrelationskoefficient som kan variera mellan 0 och 1. Ju högre koefficienten är, desto högre grad av mätprecision har skalan.

En typ av reliabilitet avser skalans **homogenitet**, det vill säga i vilken utsträckning item i en skala mäter samma egenskap. Det vanligaste homogenitetsmättet är Cronbach's alpha, som bygger på medelvärdet av samtliga interitemkorrelationer i skalan. Ett annat sätt på vilket en skalas homogenitet kan uppskattas, är att dela in den i två halvkor och beräkna sambandet mellan dem, så kallad **split-half**.

I skalor som mäter prestation under stark tidsbegränsning, är det inte lämpligt att uppskatta reliabiliteten med hjälp av ovan nämnda metoder. I stället uppskattas så kallad **test-retest** reliabilitet. Det innebär att man bjuder testet till samma personer vid två i tid separata testtillfällen och därefter beräknar sambandet mellan personernas resultat vid de två tillfällena.

Parallelltestreliabilitet innebär att man bjuder två likvärdiga versioner av en skala till samma person, och beräknar ett sambandsmått mellan de två uppsättningarna poäng.

I de fall ett testresultat baseras på bedömarens skattningar, ligger oftast **interbedömarreliabiliteten** till grund för att uppskatta mätprecisionen.

Det finns ingen enskild metod som bäst uppskattar ett tests reliabilitet. Generellt kan sägas att ju fler metoder som används och redovisas för i testmanualer, desto bättre underlag har testanvändare att bedöma testets mätprecision.

Mätfel

Alla psykologiska test är behäftade med mätfel och det är viktigt att vid tolkning av testresultat få en uppfattning om hur stort detta mätfel är. Med hjälp av reliabiliteten och standardavvikelsen i en skala kan ett poängintervall kring en testpoäng fastställas, inom vilket poängen skulle falla om en person testades ett oändligt antal gånger. Poängintervallet baseras på det så kallade medelfelet (Standard Error of Measurement) och förkortas SEM.

Genom att använda nedanstående formel för SEM fastställs det poängintervall (råpoängen +/- medelfelet) inom vilket en persons "sanna poängvärde" faller med 68 % säkerhet; alltså vid två av tre testningar. För att skapa ett något bredare, men säkrare, intervall (dvs 95 % konfidensintervall) multipliceras SEM med 1.96. Ju högre reliabilitet och ju lägre standardavvikelse en skala har, desto mindre blir medelfelet.

$$s \sqrt{1 - r}$$

Validitet

Ett tests validitet berör i vilken utsträckning testet mäter det som det avser att mäta – att det är giltigt för att ge ett mått på egenskapen i fråga. Klassisk testteori räknar med olika typer av validitet.

Innehållsvaliditet bestäms genom det inledande valet av item i ett test. Utifrån den kunskap man har om den egenskap man vill mäta, väljer man item som man finner relevanta och urskiljande för egenskapen i fråga.

Begreppsvaliditet beskriver i vilken utsträckning testet mäter avsett teoretiskt begrepp eller egenskap. Egenskapen blir i testets item operationellt definierad. Statistiska metoder såsom faktoranalys kan ge stöd för denna typ av validitet.

Kriterierelaterad validitet brukar indelas i samtidig och prediktiv validitet och visar sambandet mellan testresultat och något oberoende kriteriemått.

I testmanualer presenteras ofta bevis för testets validitet enligt klassisk testteori. I allt större utsträckning får denna konkurrens från en modernare syn på validitet, där man istället talar om ackumulerade belegg som stödjer tolkningen av testresultatet för det avsedda ändamålet (APA Standards for Educational and Psychological Testing, 1999).

Standardisering

Detta begrepp syftar i vid mening på alla de åtgärder som vidtas för att få de enskilda mätningarna så jämförbara som möjligt. Här ingår till exempel uppgiftsanalys, administreringsregler samt anvisningar för poängberäkning och tolkning. Med en snävare definition brukar man säga att standardisering är den process som leder till att man kan bestämma populationsdata (medelvärde

och spridning för populationen) och sedan använda dessa data för att kunna transformera råpoäng till standardpoäng. Denna process är nödvändig då en testpoäng utan "referensramar" inte är möjlig att tolka. Referensramarna kan antingen utgöras av en normgrupp (vanligast) eller ett kriterie.

Vägen till att omvandla en persons råpoäng till poäng på någon av de vanligast förekommande standardskalorna går över en så kallad z-transformering. Z-tabeller finns i de flesta grundläggande statistikböcker och anger antalet standardavvikelser som en testpoäng avviker från medelvärdet.

$$z\text{-poäng} = \frac{(\text{individens råpoäng} - \text{medelvärdet i gruppen})}{(\text{standardavvikelsen i gruppen})}$$

Transformerade z-poäng har normalt en spännvidd mellan -3 och +3 och ett medelvärde på 0. I praktiska sammanhang överförs oftast z-poängen till någon mer användarvänlig standardskala, som till exempel stanine eller IK. Denna transformering visas i formeln nedan.

$$(\text{individens } z\text{-poäng} \times \text{standardskalans standardavvikelse}) + (\text{standardskalans medelvärde})$$

För de i figuren återgivna standardskalorna gäller

	m	s
z-poäng	0	1
T-poäng	50	10
IK-poäng	100	15
Wechslers skalpoäng	10	3
Stanine	5	2
Sten	5.5	2

Ett annat sätt att utvärdera en råpoäng är att se var i fördelningen den faller – oavsett i vilken grad fördelningen är snedfördelad. För detta ändamål används percentiler, som talar om hur stor andel av testresultaten som faller under en viss råpoäng.

