

**Introduktion til**

**ERHVERVSTEST  
og**

**Grundlæggende Psykometri**

**Kursus i Erhvervsrettet testanvendelse (Niveau I)  
Psykologi og Erhverv**

# INDHOLD

## **1. del**

Generel introduktion om test, persontræk, adfærd og resultater i jobbet

Definition; hvad er en test?

Testtyper

Konstruktion af test

Introduktion til grundlæggende statistik

Vurdering af forskellige test

Etik

## **2. del**

Det generelle skøn

Gennemsnittet

Medianen

Spredningen

Standardafvigelsen

Fordelinger

Normalfordelingen

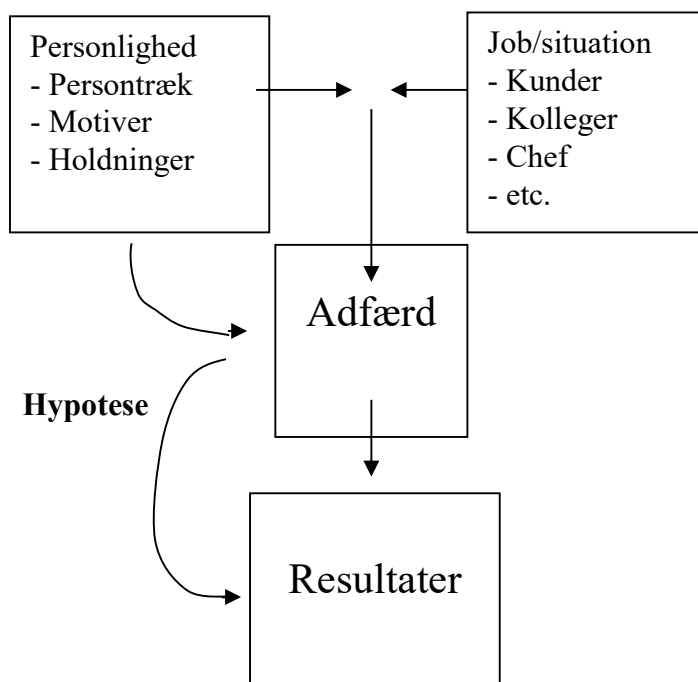
# Generel introduktion til test, persontræk, adfærd og resultater i jobbet

Inden man ansætter en person i en virksomhed, vil man gerne vide, hvilke resultater han/hun kan skabe i det aktuelle job i virksomheden.

Når man f.eks. ansætter en sælger, er det interessante, om han kan sælge noget. Det er mindre interessant, om han er varm og følsom eller udadvendt og kreativ eller dominerende eller hvilke andre træk en bestemt psykologisk test kan afdække.

Virksomheden er naturligvis interesseret i, om han passer ind i virksomhedens miljø. Han skal også behandle kunderne ordentligt, men det kan han jo gøre på mange måder.

Det stiller krav til testbrugeren om at bruge testen på en måde, som tager udgangspunkt i virksomhedens erfaringer.



## Adfærd

Gennem en psykologisk test kan vi få beskrevet et menneskes personlighedsstruktur. Det vil sige, at vi får afdækket hvilke persontræk, der er særlig fremtrædende og hvilke der er knap så fremtrædende. Men det siger stadig ikke noget om, hvordan dette menneske vil opføre sig.

En person, der er kreativ, kan i et restriktivt miljø virke useriøs og barnlig. I et miljø hvor der lægges vægt på nytænkning, kan samme person virke engageret, seriøs og ansvarsfuld.

Vi må altså tage udgangspunkt i den situation eller det miljø, hvor vi ønsker at placere personen for at komme fra en beskrivelse af persontræk til en beskrivelse af personens sandsynlige adfærd.

## **Resultater i jobbet**

Når man forudser en sandsynlig adfærd i en bestemt situation, er det en hypotese, der bygger på testresultater og analyse af situationen. En forestilling om hvilke resultater denne adfærd vil skabe for virksomheden, vil være en yderligere hypotese.

Man kan ikke dokumentere en direkte sammenhæng mellem specielle persontræk og resultater i et bestemt job. Men ud fra erfaringer med test kan der opstilles realistiske hypoteser.

Det kan ikke lade sig gøre at udvælge en tilfældig stor gruppe mennesker, teste dem med en bestemt test og lade dem alle blive f.eks. salgschefer i samme job for at se, om der er en statistisk sammenhæng mellem bestemte konstellationer af træk og succes som salgschef.

Omvendt har man testet en stor gruppe salgschefer med en bestemt test for at se, om bestemte træk går igen.

Hvis resultaterne her var entydige, kunne de naturligvis give et fingerpeg men ikke noget bevis. Men resultaterne er ikke entydige. Derfor må konklusionen være, at der er mange måder at skabe gode resultater på.

## **Brug af psykologiske test**

Ved hjælp af test får vi meget hurtigt nogle nuancer frem om personlighedsstrukturen eller om færdigheder og evner. Et indblik man ellers kun har hos mennesker, man kender meget godt.

De nuancerede beskrivelser kan give et godt grundlag for at opstille hypoteser om fremtidige resultater i forskellige omgivelser. Værdien af disse forudsigelser bygger i høj grad på testbrugerens dygtighed og erfaringer. I den optimale anvendelse af test bliver testresultaterne verificeret og forklaret i et efterfølgende interview.

Test kan anvendes ved rekruttering eller ved anden form for udvælgelse og vurdering.

Da resultaterne fra en test er udtrykt i en talskala (eller ved personlighedsprøver et kategorisystem), giver de grundlag for en præcis kommunikation. En rigtig tilbagemelding kan samtidig give større klarhed og sikkerhed hos den testede om egne stærke og svage sider.

Det er en stor fordel i forbindelse med karriereplanlægning, hvor der på baggrund af testresultatet kan formuleres nogle præcise vurderinger om evner og egenskaber.

Test er også velegnede i forbindelse med genplacering. I situationer hvor mennesker er uden job og skal søge nyt, har de en tendens til at miste selvtilliden. Samtidig udvikler de en diffus opfattelse af deres egne stærke og svage sider. Her giver "de objektive" resultater fra en test støtte til at fastholde en præcis selvsvurdering, som er meget central i en jobsøgningsfase.

Ved anvendelse af test er det imidlertid særdeles vigtigt at overveje, om testen måler det, man ønsker at måle. I forbindelse med rekruttering er det derfor vigtigt først at analysere jobbet, som ønskes besat. Hvad kræves i jobbet, og hvilke personsegenskaber er relevante i jobbet.

Det er ofte forskellige personsegenskaber, der er centrale i forskellige job.

I en anden situation er det måske vigtigt at vurdere hvilke personsegenskaber, der mangler i ledergruppen. Mangler der visioner og kreativitet i ledergruppen? Eller er der for mange visionære og kreative personer i ledergruppen?

Det er ikke sikkert, at samme test kan anvendes til at vise det hele. Derfor kan det blive nødvendigt at bruge flere test.

## **Kombination af test og interview**

Det er vigtigt at fastholde, at der ved anvendelse af test opereres med sandsynligheder. Nogle vil derfor på baggrund af testresultatet falde uden for de stillede krav, selv om de burde have ligget inden for. Omvendt vil der være nogle, der på baggrund af testen slipper gennem nåleøjet uden at besidde de relevante egenskaber og kvalifikationer.

En supplerende vurdering, f.eks. i form af et interview, vil derfor altid øge sikkerheden.

Kombinationen af test og interview er i det hele taget ideel.

Test er interviews overlegne m.h.t. statistisk validitet når evner, færdigheder og persontræk skal beskrives.

Til gengæld er interviews bedre til at forudsige fremtidige jobresultater når udgangspunktet er interesser, karriereønsker og resultater fra tidligere jobadfærd.

Personlighedstest giver desuden ofte en legitim anledning til at tale om private personlige områder, som det er vanskeligt at komme ind på i et interview.

## **Definition; hvad er en test?**

„En test er en standardiseret metode til at iagttage og beskrive et menneskes færdigheder, evner eller persontræk." (Oluf Aagaard, IP, 1989)

Der stilles to overordnede krav til en test. For det første til den måde en person iagttages på. For det andet til den metode der anvendes, når iagttagelsen beskrives.

Iagttagelsen eller undersøgelsen skal foregå under kontrollerede forhold, og beskrivelsen skal kunne omsættes til en talskala eller et kategorisystem.

Når vi i dagligdagen iagttager andre mennesker, har det samme elementer som en testsituation. Vi vurderer, sammenligner og kategoriserer andre mennesker, når vi ser dem.

Vi har en tendens til at vurdere mennesker ud fra de typiske handlemåder de har. Vi kan se, at en bestemt person handler på samme måde i en række situationer, der ligner hinanden. Og vi kan se, at andre mennesker handler anderledes i disse situationer.

Disse forskelle og ligheder får os til at tale om færdigheder, evner eller persontræk, der er karakteristiske for det enkelte menneske.

Problemerne med iagttagelserne i dagligdagen er blot, at vi sjældent oplever mennesker i helt identiske situationer. Derfor bliver sammenligningerne ikke helt præcise.

Situationer, der forekommer ens, kan ofte være meget forskellige. De kan være sammensat af uoverskueligt mange forhold, der påvirker en persons adfærd. Samtidig er vor oplevelse af situationen subjektiv. Den er styret af vort eget personlige sammenligningsgrundlag og måde at kategorisere på. Det betyder, at vore beskrivelser af samme iagttagelse kan blive forskellige. Når vi skal beskrive og kommunikere forskelle og ligheder mellem mennesker, så de bliver forstået på samme måde af alle involverede, er to ting afgørende:

- Vi skal have en veldefineret situation (testsituationen).
- Vi skal have et entydigt system til at beskrive forskellene og lighederne (eksempelvis en talskala eller et kategorisystem).

## Testtyper

Ved valg eller konstruktion af en test skal testen og testsituationen ligne den situation, vi skal forudsige noget om. Jo større lighed der er mellem testsituationen og den virkelige situation, jo større er mulighederne for at kunne forudsige noget relevant på baggrund af iagttagelserne i testsituationen.

De erhvervspsykologiske test kan bl.a. grupperes på følgende måde:

1. Jobsimulerede test.
2. Færdighedsprøver.
3. Intelligenstest og omtankeprøver.
4. Personlighedstest.

Denne rækkefølge afspejler graden af lighed mellem testsituationen og den fremtidige situation, som vi vil sige noget om. De jobsimulerede har den største lighed.

## Jobsimulerede test

Ved jobsimulerede test er der tale om test, der har deres udgangspunkt i en nøje analyse af et bestemt job. Særligt kritiske variabler eller situationer fra jobbet bliver konstrueret som opgaver i en test.

Ingen test kan imidlertid forudsige noget om fremtidig adfærd og jobsucces med 100% sikkerhed. Enhver testsituation vil være begrænset i forhold til virkelighedens mangfoldighed og evige forandring.

Alligevel kan man gennem den adfærd, man registrerer under testen få et fingerpeg om, hvordan testpersonerne vil reagere i fremtidige virkelige situationer.

Ved udvælgelse af piloter anvendes eksempelvis et simuleret cockpit, hvor alt udstyr er identisk med udstyret i et rigtigt fly. I dette cockpit kan kandidaterne sættes i kritiske situationer, der ligner virkelige situationer.

Der kan udarbejdes et tal på, hvor meget testen bidrager til en bedre udvælgelse, ved at følge de udvalgte i flere år og se hvem der får succes. Hvis den traditionelle udvælgelsesmetode kombineres med resultaterne fra testen, bliver sandsynligheden for succes eksempelvis øget med omkring 10%, hvilket kunne være fra ca. 60% til ca. 70%.

Men der er stadig ingen absolut sikkerhed for, at de rigtige mennesker udvælges til piloter.

## **Færdighedsprøver**

Ved færdighedsprøver måler man de tillærte eller iboende færdigheder, som prøvetageren besidder på prøvetidspunktet, og som han/hun kan få brug for i en kommende jobsituation eller uddannelsessituation.

Prøverne går ud på at iagttage adfærden, eller hvor god kandidaten er til at løse de stillede opgaver.

Det drejer sig ikke om at afdække den personlighed, der ligger bag ved adfærden men selve den medfødte eller tillærte adfærd. Selv om prøven ikke er rettet mod et bestemt job, nærmer vi os en mulighed for at forudsige, hvordan personens resultater i jobbet vil være.

Det drejer sig bl.a. om:

- Sansetest
- Motoriske test
- Staveprøver
- Sprogtest
- Vidensprøver
- Regneprøver
- Grundlæggende færdigheder

Ved nogle af disse prøver er der tale om at måle relativt grundlæggende færdigheder.

Med sansetest undersøges sansorganernes effektivitet, eksempelvis synsprøver, høreprøver etc.

Med motoriske prøver undersøges muskelbevægelsernes hastighed og præcision, f.eks. reaktionstidsprøver, fingerfærdighedsprøver og koordinationsprøver.

Prøver, der måler tillærte færdigheder, minder om de prøver, de fleste af os har været ude for ved eksaminer eller terminsprøver i skolen. Princippet i staveprøver, sprogtest, vidensprøver og regneprøver er i hovedtræk det samme.

Ved prøven sættes en gruppe personer til at løse en række opgaver, og det konstateres, at nogle løser flere opgaver end andre.

Det er de individuelle forskelle, der er af interesse. Man ønsker at iagttage, hvem der er bedre end andre.

I praksis anvendes prøven til at registrere, hvor mange opgaver en enkelt person løser i forhold til "landsgennemsnittet" for personer med en tilsvarende uddannelse eller fra samme målgruppe.

Eller man vurderer, om de færdigheder en bestemt person besidder, kan honorere de færdigheder, et konkret job kræver. Dette gøres på baggrund af erfaringer med det konkrete job eller på baggrund af en analyse af det niveau, der er nødvendigt i jobbet.

## **Fem Faktor Modellen**

Den nyeste forskning inden for personlighedspsykologien bygger dog primært på Fem Faktor Modellen eller The Big Five (Digman 1990, Wiggins 1996). Som bekendt er modellens udgangspunkt, at personligheden kan beskrives ved fem brede og grundlæggende personlighedstræk: Neuroticisme, Ekstroversion, Åbenhed, Venlighed og Samvittighedsfuldhed. Traditionen bygger på den såkaldte leksikon model, der kan spores tilbage til Allport (Allport 1936) og Cattell (Cattell 1965). Antagelsen er, at dominerende personlighedsdimensioner vil afspejle sig i dagligdagssproget, og at man kan få en ide om personlighedens struktur ved at analysere personlighedsbeskrivende adjektiver i sproget.

Cattell kom på baggrund af faktoranalyser frem til 16 faktorer og hævdede, at man med disse 16 faktorer kunne beskrive personligheden udtømmende (Cattell et al 1970, Cattell 1989). Senere faktoranalyser med mere moderne teknologi ender med 5 faktorer (McCrae & Costa 1996).

Eysenck har en anden tilgang, han ville ud over at beskrive personligheden, også forklare de individuelle forskelle, han observerede. Denne tradition inddrager således i højere grad data og teorier fra andre områder af psykologien, herunder eksperimentalpsykologi, neuropsykologi og adfærdsgenetik (Eysenck & Eysenck 1985). Eysenck opererer kun med fire faktorer, men en af faktorerne kan med stor ret deles op og således danne Fem Faktor Modellen.

Fem Faktor Modellen kan betragtes som en syntese af forrige århundredes forskning, og modellen er blevet så dominerende, at en meget væsentlig del af den empiriske personlighedsforskning enten tager direkte udgangspunkt i modellen eller i det mindste relateres indirekte til modellen. De begreber man finder i Fem Faktor Modellen og de facetter, der knytter sig til de overordnede træk, går da også igen i langt de fleste test på markedet.



## Intelligenstest og omtankeprøver

Debatten om intelligensbegrebet har været meget følsom i hele sidste århundrede, og der er endnu ikke kommet nogen afklaret eller pragmatisk holdning til begrebet. Tværtimod er der kommet flere forskellige definitioner ind i billedet eksempelvis emotionel intelligens (som vi vil diskutere nærmere på kurset).

På trods af at begrebet ikke er særlig præcist afklaret, har det altid været meget prestigefyldt at være intelligent eller at score højt på intelligensprøver.

Det hænger blandt andet sammen med, at intelligenstest har været anvendt til at udskille åndssvage og debile personer fra "normalbefolkningen", og at intelligenstest har været anvendt til at forudsige succes i skolesystemerne.

Samtidig har der været en forestilling om, at intelligens er en uforanderlig del af det enkelte menneske.

Efterhånden er der dog opstået en mere nuanceret forestilling om intelligens. For det første at intelligens ikke er uforanderlig. For det andet at intelligens er af en meget kompleks natur, som kan komme til udtryk på forskellige måder hos forskellige mennesker.

Et generelt udtryk for intelligens kan være, hvor godt en person klarer sig i tilværelsen.

Et andet mere snævert udtryk for intelligens er, hvor godt en person kan løse komplicerede opgaver.

Ofte er der god sammenhæng mellem disse to udtryk, men ikke altid.

Intelligenstest og omtankeprøver tager ofte udgangspunkt i opgaver af mere eller mindre kompliceret karakter. Dermed er de med til at indsnævre intelligensbegrebet.

I kliniske sammenhænge forsøger man at råde bod på dette ved at anvende meget sammensatte test med mange forskellige slags opgaver.

De intelligenstest, der anvendes inden for erhvervspsykologien, er dog mindre ambitiøse. Næsten alle indeholder opgaver, hvis løsning kræver:

## Omtanke og analytiske evner

*Eksempler:*

1) *Hvad er det næste tal i rækken ? 8 9 11 12 15 16 ?*

2) *Alle englændere kan lide te. Jenny kan ikke lide te. Jenny er englænder.*

*- Gå ud fra, at de to første udsagn er rigtige.*

Er det sidste udsagn da: 1. rigtigt. 2. forkert. 3. hverken rigtigt eller forkert.

Foruden opgaver af ovennævnte type kan der forekomme opgaver, hvis løsning kræver:

## Almen viden

Eksempel:

Hvad hedder hovedstaden i USA?

New York  Los Angeles  Washington

## Talforståelse

Eksempel:

A og B skal dele 12 kr., således at A får dobbelt så meget som B. Hvor meget får de hver?

## Sproglig forståelse

Eksempel:

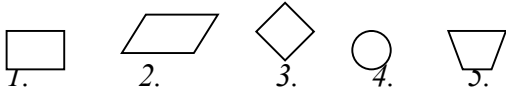
Hvilke to af nedenstående ord har sammen betydning?

1. rar 2. sensibel 3. sjælden 4. venlig 5. vittig

## Rumopfattelse

Eksempel:

Hvilken af nedenstående fem figurer passer ikke ind i sammenhængen?



## Ordforråd

Eksempel:

Sæt kryds ved det af de fire ord, der betyder det samme som ordet "Rimsmed"

Grovsmed  Digter  
 Forlover  Tågebanke

## Opmærksomhed

*Eksempel:*

*Hvor mange af nedenstående talpar er helt ens?*

19674 -19674

837191 - 839171

1596438 -1596438

97154632-97156432

## Særlige problemer ved tolkningen af resultater fra en intelligens test

En umiddelbar tolkning er:

*„En person, der er god til at løse de opgaver og problemer, som findes i intelligens testen, er som regel også god til at løse dagligdagens problemer!“*

Men:

En person kan i forbindelse med dagligdagens problemer være begrænset af :

- Personlig usikkerhed, mindreværd etc.
- Manglende fantasi og kreativitet.

En person, der ikke kan løse opgaverne i intelligens testen, kan:

- Have gode praktiske erfaringer og færdigheder.
- Have forståelse for politiske og sociale sammenhænge.

I en intelligens test er opgaverne udvalgt på en sådan måde, at løsningen ikke kræver en speciel uddannelse, men er udtryk for en generel evne. I praksis er det dog vanskeligt, idet de fleste længerevarende uddannelser indeholder uddannelse i analytiske metoder både inden for talbehandling og sprogbehandling.

Da der samtidig er stor forskel på, i hvilken grad dette formål er tilgodeset i de eksisterende test, er det vigtigt at der findes normer for de forskellige uddannelsesniveauer.

## Personlighedstest

De fleste personlighedstest afdækker særligt udvalgte personlighedstræk hos testpersonen.

Eksempelvis:

- Selvsikkerhed.
- Ansvarlighed.
- Indadvendt/udadvendt.

- Kreativitet.
- Fleksibilitet.
- Detailorientering.

En enkelt test dækker typisk fra 4 til 24 træk, men selv en test med 100 forskellige træk ville ikke kunne afsløre hele personligheden. En personlighed er så nuanceret og unik, at det praktisk taget er umuligt at beskrive den i sin helhed. Derfor er det nødvendigt, at vi udvælger særlige områder af personligheden, som er interessante for vores formål.

Et andet problem ved at vælge for mange træk er, at resultatet bliver uoverskueligt. I test hvor der eksempelvis er valgt mere end 16 træk, har det vist sig nødvendigt at gruppere trækkene for at få et overblik.

På den anden side giver test med kun 4 træk ikke noget særligt nuanceret billede af personen. Til gengæld tilgodeser test med 4 træk vores behov for enkle kategorier, og er særligt velegnede til pædagogiske formål.

I de store, gode test er der 12-16 træk, som hver især fremkommer af mindst 10-20 forskellige spørgsmål. 12-16 træk kan give så mange kombinationsmuligheder, at de kan give et særdeles nuanceret billede. En yderligere udvidelse ville derimod gøre resultatet umuligt at kommunikere og overskue.

### **Følgende spørgsmål er hentet fra forskellige test på det danske testmarked:**

1. Af de to udsagn nedenfor skal et vælges.

- *Jeg er i fysisk vigør.*
- *Jeg vil gerne have succes.*

2. Ud af de tre påstande nedenfor skal der svares; passer mest/mindst til to af dem.

- *At have frihed til at gøre som jeg har lyst til.*
- *At andre er enig med mig.*
- *Også at stå mig godt med de knap så succesfulde.*

3. Af de fire påstande nedenfor skal to besvares; et med enig, et med uenig.

- *Er ligeglad med, hvad andre mener om mig.*
- *Stræber meget for at nå de mål, jeg sætter mig.*
- *Er ikke interesseret i at vide, hvorfor andre tænker og handler, som de gør.*
- *Ønsker at have retten til at bestemme i den sidste ende.*

4. Et af følgende udsagn skal vælges.

- *Saglig*
- *Udførlig*

5. I relation til det følgende udsagn skal man vurdere sig selv på en skala fra 1-10.

- *Overbevisende - kan lide at sælge, kan lide at ændre andres mening, argumenterer overbevisende.*

6. I relation til det følgende udsagn skal der svares: Passer godt, passer nogle gange, eller passer dårligt.

*„Kommer du med tankeløse bemærkninger eller beskyldninger, som du senere fortryder?“*

7. Ud af de fire nedenstående udsagn skal der svares mest/mindst ved to af dem.

- *Mild*
- *Overtalende*
- *Beskeden*
- *Original.*

Ovennævnte spørgsmål repræsenterer to forskellige spørgeteknikker, som begge har vist sig velegnede.

Det drejer sig om ”tvunget valg metoden” og ”skalametoden”.

Ved begge metoder er det personen, der beskriver sig selv. Det er derfor afgørende, at han/hun kender sig selv og svarer ærligt.

Ganske vist er det relativt nemt at gennemskue, hvis en person har svaret uærligt. Men hvis en person svarer uærligt, er resultatet vanskeligt at anvende.

Hvilke persontræk, der er valgt til den enkelte test, er baseret på forskellige modeller af verden.

Hogrefe Psykologisk Forlag udgiver bl.a. persontesten NEO PI-R, der bygger på Femfaktormodellen eller ”The BIG FIVE”. Udgangspunktet er, at personligheden kan beskrives ved fem brede grundlæggende personlighedstræk, som er skabt gennem en syntese af forrige århundredes amerikansk og britisk personlighedsforskning. Den amerikanske tradition bygger på den såkaldte leksikonmodel, hvor man finder trækkene ved en faktoranalyse af alle personlighedsbeskrivende adjektiver i sproget. Styrken ved modellen er, at den ved faktoranalyse genfindes i mange forskellige datasæt fra andre personlighedstest.

I andre test er valget baseret på forskellige psykologiske modeller. Det er naturligvis vigtigt, om de valgte træk kan give en bred og relevant beskrivelse af personlighedsstrukturen. Men om det er den ene eller den anden nuance af et træk, der beskrives, er ikke afgørende for testens anvendelighed.

Det afgørende er, om man i virksomheden kan anvende testens kategorier.

Når ordene fra et testsystem er blevet en del af ledelsessproget i virksomheden, er det meget vanskeligt at skifte til en anden test med andre ord og beskrivelser.

## **Projektive test**

Projektive test er en anden form for personlighedstest, hvor personen fuldender påbegyndte sætninger såsom:

*Jeg er ofte...*

Eller beskriver hvad der sker i diffuse billeder med personer i forskellige situationer (TAT, Thematic Apperception Test), eller hvad vedkommende ser i nogle særligt udvalgte blækklatter (Rorschach).

Ved at svare på noget diffust og uafsluttet kommer personen til at give udtryk for noget i sig selv frem for at forholde sig til det konkrete han ser.

Det kræver dog overordentlig stor træning og psykologisk baggrund at tolke disse test. De anvendes ikke, og bør heller ikke anvendes, i forbindelse med rekruttering og karriereplanlægning.

## **Begrænsninger ved personlighedstest**

Problemerne ved personlighedstestene er, at de personlighedstræk, som bliver afsløret, ikke siger noget om, hvordan en person handler i bestemte situationer.

Samme persontræk kan give anledning til forskellig adfærd, og en bestemt adfærd kan være udtryk for forskellige persontræk.

Det vil altid være op til testbrugeren at afklare, hvordan bestemte kombinationer af træk kommer til udtryk i adfærden i forskellige situationer.

Og det vil derfor altid være nødvendigt at kombinere en personlighedstest med interviews, hvis man vil gøre sig håb om, at kunne forudsige noget sandsynligt om fremtiden.

## **Konstruktion af test**

Når formålet med prøven er fastlagt, og målgruppen indkredset, udvælges opgaverne.

Opgaverne eller svarmulighederne skal være relevante ud fra formålet med testen. Og de skal have en sværhedsgrad eller en karakter, så de kan afdække forskellene blandt målgruppens medlemmer.

Det gælder om at vurdere, hvordan en person vil klare sig i forhold til andre i en konkret fremtidig situation.

## Valg af opgaver (items) til en test

Da en prøves primære formål er at konstatere forskelle og ligheder blandt prøvetagerne, har en opgave, som alle løser rigtigt ingen værdi. Den differentierer ikke blandt prøvetagerne. Det samme gør sig gældende for en opgave, som ingen af prøvetagerne kan løse.

For at en opgave skal have nogen værdi i forsøget på at se forskelle og ligheder, skal mindst 20% af prøvetagerne kunne løse opgaven eller vælge svarmuligheden. Samtidig skal højst 80% kunne løse opgaven eller vælge svarmuligheden.

Ved konstruktion af f.eks. en regneprøve kan man gå frem på to måder. Enten vælger man en række opgaver med samme sværhedsgrad svarende til det niveau, man ønsker at undersøge. Eller man konstruerer en prøve, hvor opgaverne bliver vanskeligere, efterhånden som man kommer frem i prøven.

Ved den første prøve er tidsfaktoren afgørende. Instruktionen til prøvetager bliver at løse så mange opgaver som muligt inden for et afgrænset tidsrum (typisk 10-15 min.).

Ved den anden prøvetype er tidsfaktoren ikke afgørende. En person kommer til et vist punkt i prøven, og kan ikke løse flere af opgaverne, uanset hvor lang tid vedkommende får. Denne metode er især velegnet ved intelligensprøver.

I praksis er de fleste regneprøver og intelligensprøver dog konstrueret som en kombination af de to metoder. Tidsfaktoren bliver derved en variabel, som kan gøre prøven lettere eller vanskeligere.

## Vurdering af forskellige test

Der er en række spørgsmål, som må besvares, når en virksomhed tager stilling til brug af test.

- Passer testen til virksomhedens formål?
- Hvilke grupper er der udarbejdet statistik på?
- Måler testen præcist?
- Er begreberne i testen nøje beskrevet og afprøvet (i Danmark)?
- Er den let at administrere?
- Bliver den vedligeholdt?
- Er der løbende opfølgning på erfaringerne?
- Prisen!!!

I en norsk undersøgelse fra 1990, foretaget af Dansk Management Forum/IP's søsterorganisation NIPA, har man forsøgt at give de mest kendte personlighedstest karakter. Man har vurderet dem i forhold til, hvor godt de kan dokumentere, at de lever op til de forskellige kvalitetskrav. (Det er stort set de samme test, der findes på det danske marked). Enkelte af disse test kommer ud som relativt uegnede, idet de mangler en statistisk bearbejdet fordeling. Andre er uegnede, fordi begreberne er direkte oversat fra engelsk uden en egentlig efterprøvning i den lokale norske kultur. Det er dog de få. Næsten alle de kendte test lever pænt op til kvalitetskravene. Og det er kun nuancer, der adskiller disse test kvalitativt. Et særligt vigtigt begreb ved vurderingen af en test er testens validitet. Måler testen det, den skal måle?

For personlighedstest er det, som omtalt tidligere, vanskeligt at påvise en entydig sammenhæng mellem resultatet i testen og de fremtidige resultater i et konkret job. Enkelte testudbydere har dog tal på validiteten af den test, som de udbyder. Men disse tal kan være udtryk for mange ting og skal behandles med stor varsomhed. Det vigtigste er at forholde sig til, om begreberne i testen er afprøvet og velbeskrevet i den kultur, hvor testen skal anvendes. Desuden skal de andre kvalitetsfaktorer være i orden. Kvalifikationerne hos de mennesker, der bruger personlighedstesten, er i den sidste ende afgørende for, om resultatet af testen kan anvendes. - Har disse mennesker psykologisk forståelse og viden? - Har disse mennesker livs- og erhvervs erfaring? - Og har disse mennesker ærbødighed over for det rent hypotetiske i testresultater?

## **Etik**

Den enkelte prøvetager opfatter ofte resultatet af en psykologisk test som noget meget personligt og privat.

Det er derfor vigtigt, at oplysningerne bliver behandlet med omtanke og respekt efter følgende retningslinier:

1. Den person, der bliver udsat for en prøve (prøvetager/testperson) bør informeres om formålet og om, at resultatet ikke i sig selv er afgørende, men vil indgå i en større sammenhæng.
2. Prøvetager/testperson bør have ret til at trække sig ud af processen og forlange prøveresultatet destrueret.
3. Prøvetager/testperson bør informeres om resultaterne og have lejlighed til at diskutere dem under en samtale.
4. Brugeren er den person, som anvender prøven og administrerer resultatet. Han/hun bør besidde sådanne menneskelige kvalifikationer, at vedkommende er i stand til at se andre menneskers forskelligheder og forstå disses muligheder.
5. Brugeren skal være uddannet i brugen af det testsystem, som anvendes. Prøveresultatet videregives ikke til andre personer eller virksomheder, men er et fortroligt anliggende mellem prøvetager og bruger og evt. øvrige ansvarlige for den proces, som prøvetager er inde i.
6. Brugeren skal sikre diskretion. Ligeledes skal det sikres, at registerlovens anvisninger om opbevaring af personlige data overholdes, og at resultaterne kun bruges til det angivne og aftalte formål.



## Introduktion til grundlæggende statistik

Inden for måling af psykologiske fænomener, hvilket benævnes psykometri, anvendes statistik i stor udstrækning, og for den kompetente testbruger er kendskab til grundlæggende statistik en nødvendighed. Statistik hjælper os med at:

- Sammenligne testscore inden for en gruppe individer på en standardiseret og meningsfuld måde.
- Undersøge sammenhængen mellem forskellige testscore.
- Beregne graden af den troværdighed vi kan tillægge en testscore.
- Undersøge på en sammenlignelig måde om en test måler det, den giver sig ud for at måle.
- Undersøge om de indsamlede og anvendte data har den rette sammensætning.

I det følgende gives en kort introduktion til nogle af de statistiske begreber, som du vil møde på kurset i Erhvervsrettet testanvendelse (Niveau I). Du har måske en forestilling om, at statistik er forbeholdt nørder og andre med en særlig matematisk udviklet hjerne, men det er absolut ikke tilfældet. Du vil blive overrasket over hvor simple de fleste statistiske begreber er, og hvor meget ekstra mening det giver at kende dem. De der har haft en blok med statistik på Universitetet eller er stødt på statistik i andre sammenhænge vil sandsynligvis nikke genkendende til de fleste begreber. Andre kan måske finde denne introduktion vanskelig og for jer gælder det, at I må notere problemstillingerne ned, så vi kan gennemgå dem på kurset.

De følgende sider omhandler primært tre forhold:

- Måling af tendensen eller det generelle skøn
- Måling af variationer
- Populationer og normgrupper

### Måling af tendensen eller det generelle skøn.

- *Hvor ligger "landsgennemsnittet" eller middelscoren ?*

Når vi taler om mål for tendensen taler vi om måder hvorpå vi kan beregne gennemsnittet af et givent antal score. For eksempel:

Hvis vi beder 10 studenter om at gennemføre en færdighedstest med 10 opgaver, hvor resultatet enten er rigtigt eller forkert, kan vi se at maksimum (**max.**) scoren må være 10 (d.v.s. alle opgaver er løst rigtigt), og at minimum (**min.**) scoren er 0. Vi kunne have samlet følgende data fra vores 10 studenter:

Skema 1. Færdighedsscore fra 10 studenter

Student nr.	1	2	3	4	5	6	7	8	9	10
Total score	6	9	3	1	4	5	7	3	5	3

Ved at beregne den gennemsnitlige score fra dette sæt data, kan vi sammenligne resultatet fra vores gruppe af studenter med andre grupper. Denne metode er både meget almindelig og meget anvendelig, når vi vil sammenligne to grupper. Hvis jeg for eksempel vil se, hvor godt min gruppe på 10 studenter klarer sig i forhold til en anden gruppe på 50, som jeg tidligere har givet denne test, vil sammenligningen af de to gruppers gennemsnit være et godt sted at begynde. Gennemsnittet tillader os at sige noget generelt om de score vi har indsamlet i den specifikke gruppe.

Før vi går videre med diskussionen om, hvordan vi beregner den centrale tendens (eller gennemsnittet), skal vi definere nogle symboler, som almindeligvis anvendes ved disse beregninger:

**X** refererer til **råscoren**, som fremkommer ved, at man sammenlægger scoren fra alle de enkelte opgaver (item) i testen, som et individ har opnået. I det ovenstående eksempel er **råscoren** det antal korrekte svar som hver student har opnået i færdighedstesten.

**N** refererer til **antallet af respondenter** eller personer i gruppen, som testes. Det kan også vises som 'n'.

$\Sigma$  anvendes til at beskrive **'summen af'**, eksempelvis  $\Sigma \frac{60}{30} = 2$

Med reference til vores gruppe med 10 studenter er  $\Sigma X = 46$ ;  
 Det betyder at **summen af alle råscore** er 46  
 (eller:  $6+9+3+1+4+5+7+3+5+3 = 46$ )

Hvis vi bruger eksemplet oven over, kan vi beregne den gennemsnitlige score, som er opnået af de 10 studenter, på tre måder:

## 1. Gennemsnittet

**Gennemsnittet** beskrives med symbolet  $\bar{X}$  og kan udtrykkes i følgende ligning:

$$\bar{X} = \frac{\Sigma X}{N} \quad \text{eller som i vores eksempel:} \quad \bar{X} = \frac{46}{10} = 4,6$$

Derfor er  $\bar{X} = 4,6$

Nu giver vi den samme færdighedstest til 15 ledere og får følgende resultat:

<b>Leder Nr.</b>	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
<b>Total score</b>	8	6	3	7	2	5	6	2	9	6	8	4	10	7	6

### Opgave 1:

Beregn gennemsnittet eller middelværdien for gruppen af ledere ved at anvende formlen:  $\bar{X} =$

$$\frac{\sum X}{N}$$

Hvem har klaret sig bedst; lederne eller studenterne?

## 2. Medianen

Medianen er et andet mål for det centrale skøn. Medianen er den midterste score, når alle score er sat op i en række fra lavest til højest. Det er den værdi, hvor 50% af de opnåede score falder under og 50% kommer over. I studentergruppen kan vi beregne værdien af medianen ved at stille alle de observerede scores op i en række, hvor vi begynder med de laveste score.

1 3 3 3 ↑ 4 5 5 6 7 9

NB! Hvis en score forekommer mere end en gang i vores data (som f.eks. "3", der forekommer 3 gange i vores data fra studenterne), skal det også vises, når vi stiller dataene op i rækken (i eksemplet oven over ville det ikke have været nok at sætte et 3-tal ind).

Da vores data består af et lige antal score, ligger middelværdien i dette eksempel mellem to værdier. Disse to værdier er forskellige, så vi bliver nødt til at beregne middelscoren imellem disse to. I dette eksempel er de to score 4 og 5; **medianen** er derfor  $4\frac{1}{2}$  eller 4,5.

Hvis midterpunktet ligger mellem to værdier, som er de samme, så er det denne værdi, der er medianen. Hvis midterpunktet falder lige præcist på en værdi, så er det også denne værdi, der er medianen.

### Opgave 2:

Hvad er medianen for vores ledergruppe?

Hvilken værdi er højest, medianen eller middelværdien?

## 3. Modaltallet

Det er den score, som er opnået af de fleste. Det er det mest almindelige resultat i et datasæt. I vores data fra studenterne er modaltallet 3.

### Opgave 3:

Hvad er modaltallet for vores ledergruppe?

Den gennemsnitlige dansker får 1,9 barn. Hvilken af de 3 metoder til at beregne dette generelle skøn er brugt?

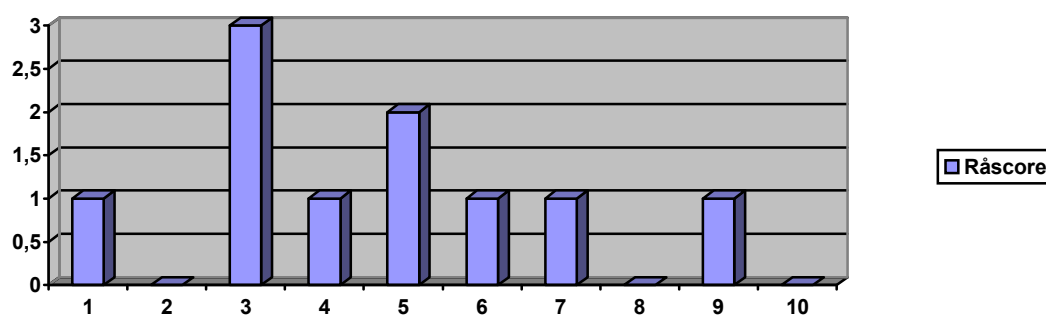
Hvordan forvirrer denne metode? Findes der en anden mindre forvirrende metode?

### **Mål for variationen. Hvor spredte er vores data?**

Når vi kender scorene for det generelle skøn på færdighedstesten for vores studenter og ledere, kan vi foretage en brugbar sammenligning mellem de to grupper. Vi kan også sammenligne vores to grupper ved at se på spredningen af scorene fra færdighedstestene. Hvis der er stor forskel på de score som studenterne har opnået, kan vi sige at denne gruppe udviser høj variation. Hvis der derimod er en lille spredning mellem de score, der er opnået i denne gruppe, er der en lav variation. Hvis vi undersøger variationen i de score, som er opnået i vores gruppe af studenter, kan vi undersøge om de har opnået meget ens resultater eller meget forskellige resultater.

En **frekvensfordeling** kan afbildes i en graf på en sådan måde, at vi grafisk kan se, hvordan vores data varierer. Vi kan eksempelvis præsentere studenternes score på følgende måde:

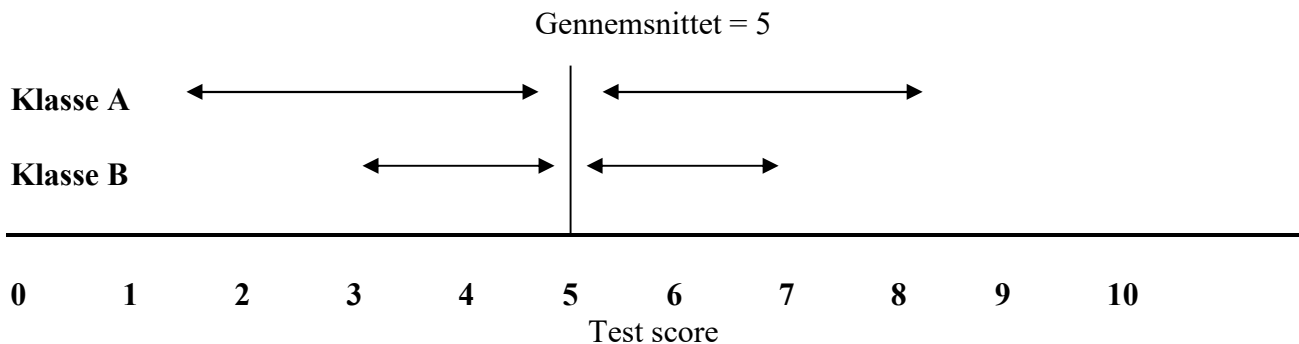
**Fig. 1.** En frekvensfordeling af studenters færdigheder



**På grafen kan vi se, at 3 studenter opnåede en score på 3 og 2 opnåede en score på 5 etc.**

At kende variationen i et datasæt giver også mening, når vi vil sammenligne datasættet med andre datasæt. Hvis du f.eks. indsamlede data fra to forskellige skoleklasser, som havde fået den samme stavetest, kunne du finde, at de havde opnået det samme middeltal (se fig. 2). Vi kan imidlertid ikke slutte at stavefærdighederne i de to klasser (A og B) er identiske. Sammenlignet med B klassen kan A klassen have mange flere elever, som scorer væsentligt højere eller lavere end middeltallet. Det vil sige at A klassens stavefærdigheder udviser en større variation, da testresultaterne er langt mere spredte.

Fig. 2. En grafisk fremstilling af resultaterne fra en læsetest i to forskellige klasser (A og B).



Der findes to måder hvorpå man kan betragte spredningen:

- Bredden og
- Standardafvigelsen

### Bredden

Bredden er det mest simple mål for **fordelingen** af score og **variationen** i de opnåede score. Bredden viser forskellen mellem den højeste og laveste score i et datasæt. I det ovenstående eksempel fra stavetesten er bredden i A-klassens score 1-9 og for B-klassen er det 3-7. (Tallene fremkommer ved at se hvor pilene begynder og ender i fig. 2.)

### Standardafvigelsen

Ligesom bredden måler **standardafvigelsen** spredningen (eller variationen) af data. Den kan beskrives som den gennemsnitlige forskel mellem alle score og deres gennemsnit.

**Standardafvigelsen** svarer på spørgsmålet: I hvilken grad afviger scorerne i denne gruppe data fra gennemsnittet? Den bliver ofte udtrykt som SD (sd) eller S (s) og bliver beregnet med følgende formel.

$$SD = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

Det er ikke nødvendigt at huske denne formel (du bliver ikke spurgt om den til eksamen på Niveau I kurset). **Standardafvigelsen** er i de fleste manualer udregnet på forhånd, så du ikke selv behøver at regne den ud. Det er imidlertid vigtigt, at du ved hvad beregningen betyder, og hvad den anvendes til.

Vi kan beregne **standardafvigelsen** i en gruppe data på følgende måde:

(Datagruppe = 17 19 20 21 23)

1. **trin.** Beregn gennemsnittet.

$$\Sigma X = 17 + 19 + 20 + 21 + 23 = 100$$

$$\bar{X} = \frac{100}{5} = 20$$

2. **trin.** Find forskellen mellem scorerne og beregn kvadratet på forskellene.

$X \div \bar{X}$	$(X \div \bar{X})^2$
$17 \div 20 = \div 3$	9
$19 \div 20 = \div 1$	1
$20 \div 20 = 0$	0
$21 \div 20 = 1$	1
$23 \div 20 = 3$	9

Grunden til at vi sætter forskellen mellem scorerne i anden er, at vi vil have fjernet de negative tal.

3. **trin.** Læg alle de kvadrerede forskelle sammen.

$$\Sigma (X \div \bar{X})^2 = 9 + 1 + 0 + 1 + 9 = 20$$

4. **trin.** Divider med antallet af individer.

$$\sqrt{\frac{20}{5}} = \sqrt{4} = 2 \quad (\text{Standardafvigelsen er derfor } 2)$$

NB! Jo større SD er desto større er variationen mellem dataene.

Opgave 4:

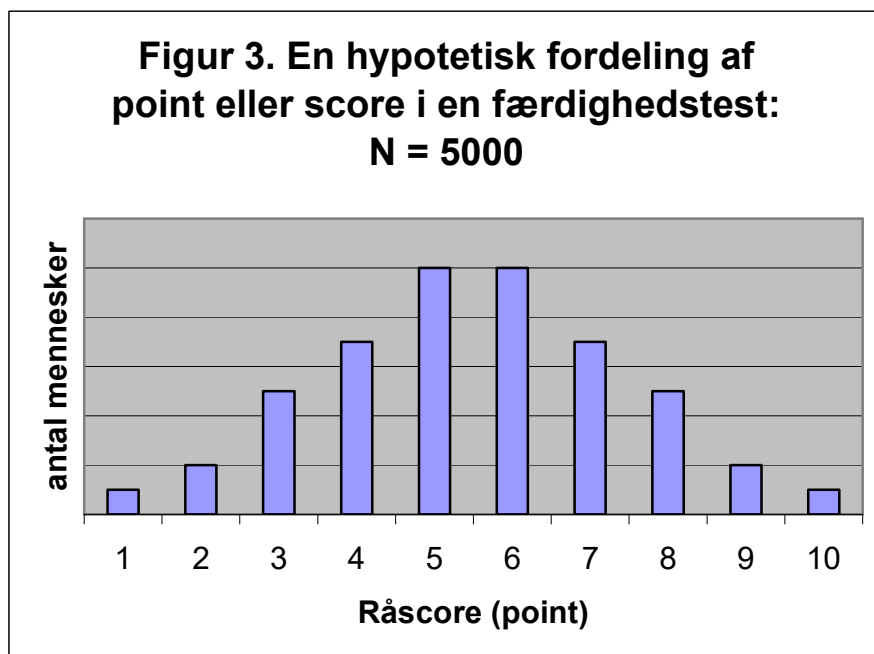
1. Beregn bredden og standardafvigelsen for studenternes score.
2. Hvad kan vi sige om variationen i studenternes score sammenlignet med eksemplet oven over?

Populationer og undersøgelseshold; hvor generelle er mine resultater ?

Et **undersøgelseshold** (de vi vælger at teste) bliver taget ud af en **population**. Populationer kan beskrives som en samling af observationer, ofte en gruppe mennesker, der deler nogle fælles karakteristika. Det er i almindelighed umuligt at måle karakteristika for hele populationer, så vi estimerer som oftest værdierne på en population ud fra værdierne fra et udvalgt undersøgelseshold. I hvilken grad scorerne fra et udvalgt undersøgelseshold repræsenterer scorerne fra hele populationen kan undersøges ved hjælp af forskellige statistiske tekniker.

**Normalfordelingen**

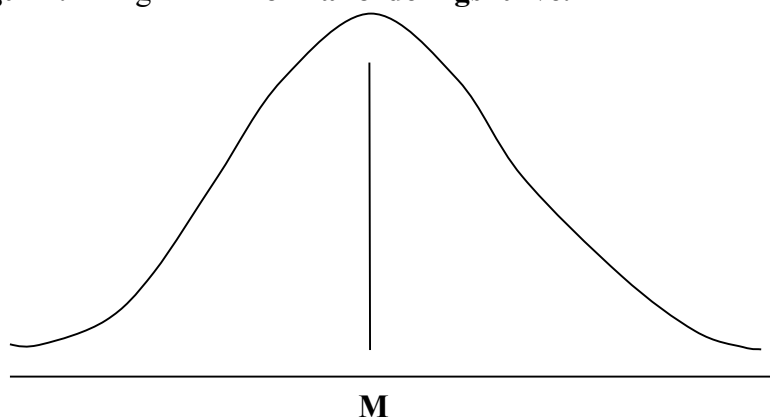
Normalfordelingen beskriver hvad vi vil forvente at en typisk fordeling af score vil ligne. Det er en teoretisk model baseret på et uendeligt antal observationer, og beskriver mange menneskelige karakteristika (eksempelvis højde, vægt og IQ). Hvis vi for eksempel havde et enormt forskningsbudget, og var i stand til at give vores færdighedstest til 5000 mennesker, ville vi have et håb om, at vores frekvensfordeling ville se nogenlunde således ud:



Vi kan se, at de fleste af vores 5000 prøvedeltagere opnåede en score tæt på midtpunktet. Som vi forventede, opnåede relativt få af vores prøvedeltagere meget høje og meget lave score. Gennemsnittet ser ud til at ligge mellem 5 og 6 fordi disse værdier fremtræder hyppigst. Det ses således, at vores fordeling af score ligner en **normalfordeling**.

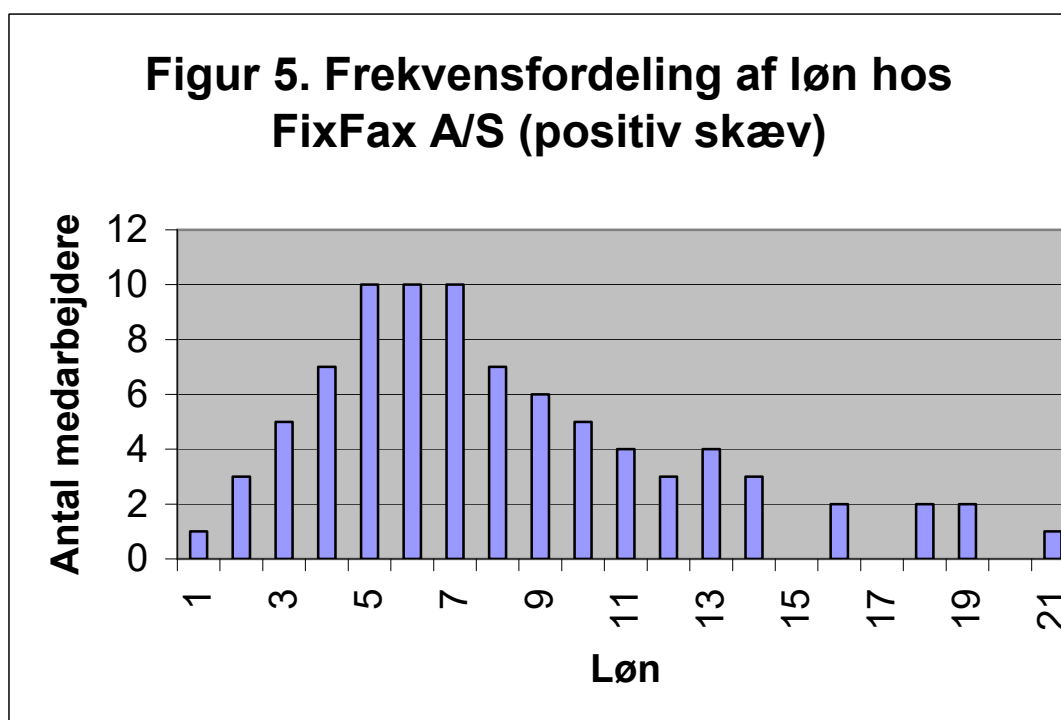
De fleste test viser fordelinger, som ligner **normalfordelingen**, når de gives til store grupper af forsøgspersoner. (Det er dog usandsynligt, at vi ser en helt eksakt gengivelse.) I sin grafiske form gengives normalfordelingen ved en klokke-formet kurve med de fleste score i midten og få ude i siderne.

**Figur 4. Den grafiske normalfordelingskurve.**



Psykometrikere søger ofte at skabe en normalfordeling, når de afprøver deres test. Når fordelingen af score i en test ligner normalfordelingen betyder det, at testen diskriminerer godt mellem høje og lave præstationer (i det mindste på den gruppe, som testen blev afprøvet på). Med andre ord; høje score og lave score opnås kun af et lille mindretal af personer.

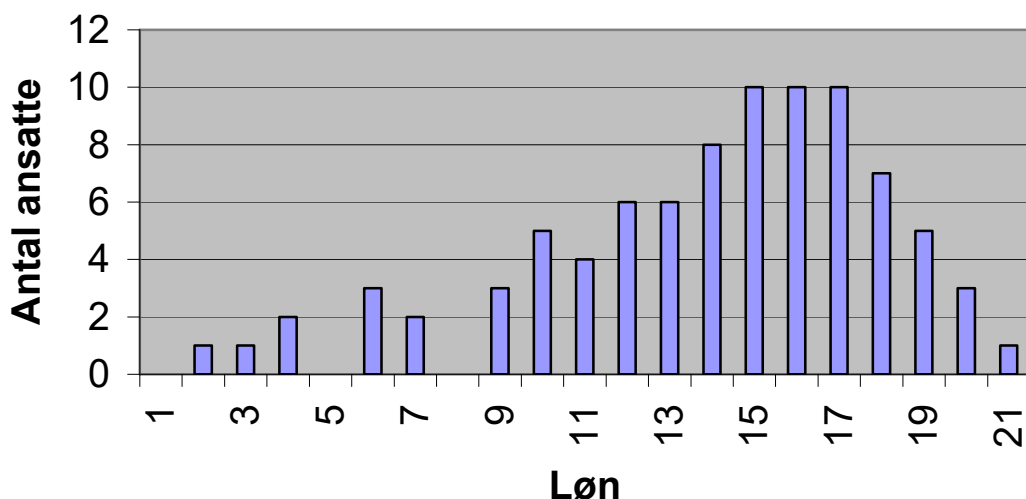
Et vigtigt karakteristika ved normalfordelingskurven er at **medianen, gennemsnittet eller det generelle skøn** ligger nøjagtigt det samme sted med 50% af resultaterne på hver sin side. Denne situation forekommer dog i praksis sjældent, men forskellen mellem medianen, gennemsnittet og det generelle skøn kan ind imellem fortælle noget vigtigt om fordelingen. For eksempel hvis vi vil udregne den gennemsnitlige indkomst blandt medarbejderne i en stor organisation (FixFax A/S) finder vi måske, at langt de fleste medarbejdere tjener nogenlunde det samme, og at nogle ganske få i topledelsen tjener væsentligt mere. I dette tilfælde vil **gennemsnittet** ligge højere end medianen og det generelle skøn. Fordelingen af indkomst siges så at være positivt skæv (se figur 5).



Hvis flertallet af ansatte i en virksomhed havde en høj løn (eksempelvis Consult A/S) og kun nogle få ansatte havde en væsentlig lavere løn vil **gennemsnittet** ligge lavere end medianen og det generelle skøn. Indkomstfordelingen siges så at være **negativt skæv** (se figur 6).



**Figur 6. Frekvensfordeling af løn hos Consult A/S (negativ skæv)**



Som det ses er hældningerne i de skæve fordelinger anderledes end i den normale fordeling. Kurvens form (især hvis den er skæv til en af siderne) kan således give os et signal om, at vores data er positivt eller negativt skæve.

*NB. Det kan være en hjælp at betragte skæve fordelinger som finner på en haj! Eks:*

*Positiv skæv*

*Negativ skæv*

*Hvis hajen svømmer fra venstre mod højre, kan vi se, at dataene er positivt skæve. Hvis hajen svømmer fra højre mod venstre, kan vi se, at dataene er negativt skævet.*

*Opgave 5:*

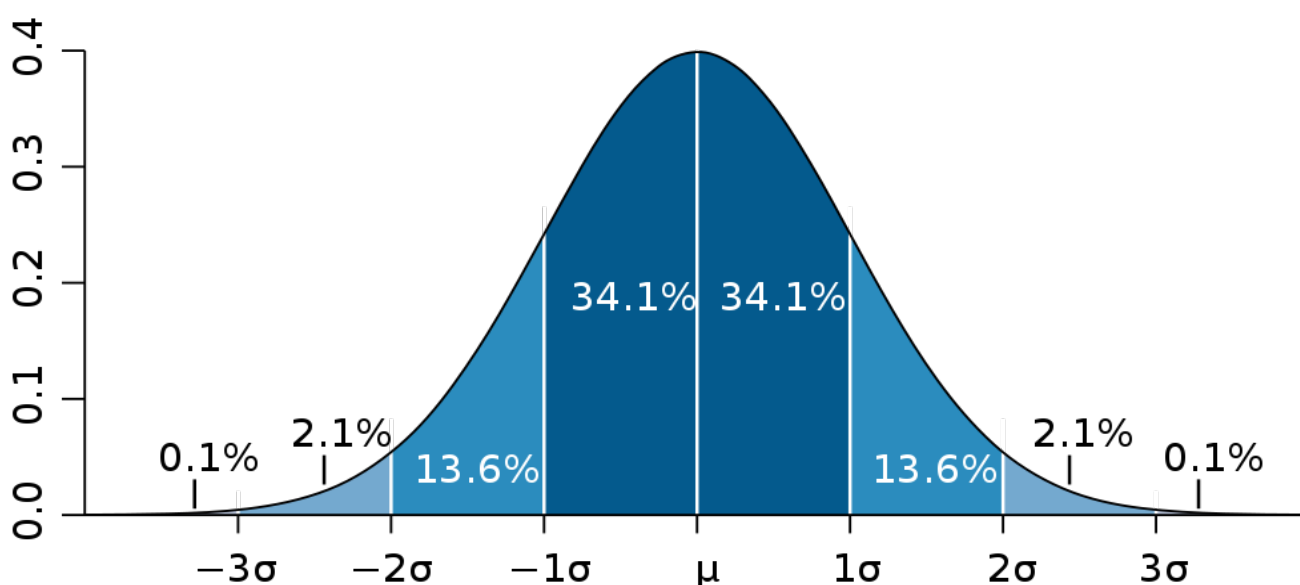
- Tegn 3 linier på graferne i figur 5 og 6 hvor du tror gennemsnit, median og det generelle skøn ligger.
- Se igen på fordelingen af resultaterne fra vores studenter i figur 1. Hvilken fordeling beskriver disse data bedst?

En perfekt normalt fordelt kurve tillader os også at forudsige hvor mange procent af resultaterne eller scorerne, der ligger inden for **én standardafvigelse** fra medianen (se fig. 7). Omkring 34% (34,13%) af resultaterne ligger mellem gennemsnittet og et punkt en standardafvigelse væk fra gennemsnittet. Det vil sige at 34% af hele arealet under kurven ligger mellem midtpunktet og en

standardafvigelse væk. Tilsvarende ligger 68% af resultaterne inden for en standardafvigelse væk fra gennemsnittet.

Næsten 48% af arealet ligger mellem gennemsnittet og et punkt 2 standardafvigelser væk (96% af resultaterne ligger inden for 2 standardafvigelser fra gennemsnittet). Omkring 49,9% af arealet (og resultaterne) ligger mellem gennemsnittet og 3 standardafvigelser væk fra gennemsnittet. Tre standardafvigelser hver vej væk fra gennemsnittet omfatter næsten alle resultaterne i en normalfordeling.

Fig. 7. Normalfordelingskurven med de procenter, der ligger i standardafvigelsen



## Resume

Dette skrift indeholder nogle af de vigtigste statistiske koncepter og begreber, som anvendes i forbindelse med erhvervspsykologisk testning. Den beskrevne statistik gør dig i stand til at sammenligne forskellige grupper af individer, når der er tale om gennemsnit eller fordelingen af resultater eller råscore og således udarbejde hypoteser om de færdigheder, der ligger i gruppen. Den beskrevne statistik gør dig i stand til at vurdere data fra en given test, og se om testen eller opgaverne i testen diskriminerer tilstrækkeligt mellem individerne.

Der bliver mulighed for at diskutere uklare punkter på kurset.