

Online kataloger med indholdsfortegnelser af samleværker II:

Undersøgelse af Danmarks Pædagogiske Bibliotek

Af Claus Poulsen og Annie Lenschau-Teglers

Abstract

Gennem 3 stikprøveundersøgelser i et middelstort dansk forskningsbiblioteks bogmasse vises, at samlingen består af i gennemsnit knap 20% samleværker. Hvert samleværk har i gennemsnit omkring 20 bidrag eller titler. Det betyder, at for hver 100 værker, der er registreret i on-line kataloget, er der ca. 400 ekstra titler, brugeren ikke kan søge på. Eller biblioteket skjuler omkring 80% af titlerne for brugeren.

Resultatet er lidt lavere end resultatet af en tilsvarende amerikansk undersøgelse, men det betyder, at Danmarks Pædagogiske Bibliotek ved at anvende 2% mere til procesomkostninger kan tilbyde en forøgelse af antallet af søgbare titler i onlinekataloget med 400%!

Hvor mange værker er ikke registreret i et online katalog?

Det er vist - på baggrund af en undersøgelse af et amerikansk unversitetsbibliotek - at der er mulighed for at 6-doble antallet af søgbare titler og derved forbedre online katalogerne således at et af de mest veldokumenterede bruger-behov tilfredsstilles (Hoffmann & Magner 1985). Det er også vist, at en sådan forbedring af online kataloget er en marginal udgift (Poulsen 1992).

Udgangspunktet for denne artikels undersøgelse er Hoffmann og Magner's undersøgelse af biblioteket på Santa Ana College Library i Californien. Dette bibliotek er et universal- og institutionsbibliotek for en en højere læreanstalt med 70.000 bind.

En stikprøve på 4.094 bøger viste at 871 var samleværker, svarende til 21.3%. Samleværkerne indeholdt i gennemsnit 31.2 titler. For hver titel i ovenstående bibliotekskatalog gemmer der sig 6 til i bogmassen. Biblioteket kan tilbyde 600% flere

titler at søge i, end i dag, **uden** at købe én bog mér. Eller: lagt op i et online katalog svarer det til at der i stedet for 70.000 er omkring 350.000 søgbare titler.

Men er det nu også tilfældet for andre biblioteker? - det ved vi intet om. Derfor har vi lavet en undersøgelse af bogmassens sammensætning på Danmarks Pædagogiske Bibliotek.

Kataloget på Danmarks Pædagogiske Bibliotek.

Danmarks Pædagogiske Bibliotek, afdelingen på Lersø Park Allé er et højt specialiseret forskningsbibliotek med ca. 350.000 monografier indenfor pædagogik og børne- og ungdomspsykologi. Den årlige tilvækst er på ca. 8.000 bøger.

Der er udtaget stikprøver på 3 forskellige måder:

Stikprøve 1: Alle bøger publiceret før 1990:

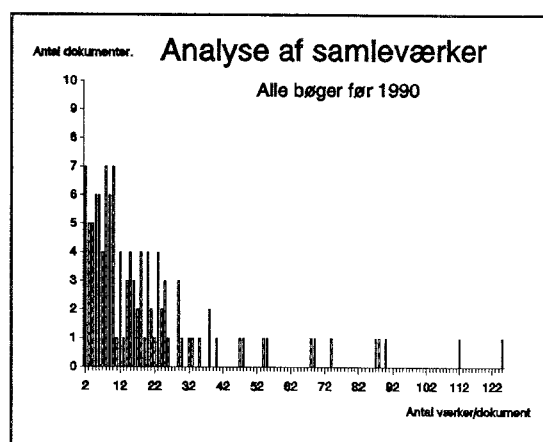
På Danmarks Pædagogiske Bibliotek er alle bøger, publiceret før 1990 opstillet for sig, da der i 1990 indførtes katalogbrud.

Vi valgte sidste bog fra 4' og 5' hylde på hver reol og udtog herved 718 bind. Antallet af værker blev noteret for hvert bind.

Prøven indeholdt 114 samleværker - dvs. samleværkerne udgjorde 16% af prøvens bind. På dette grundlag kan vi med 95% procent sikkerhed sige at der er $16 \pm 3\%$ samleværker i hele samlingen. (Se Appendiks).

Samleværkernes bidrag er vist på Figur 1.

Vi kan på dette grundlag sige, at det gennemsnitlige antal værker pr. samleværk i hele samlingen er 20 ± 4 med 95% sikkerhed. (Se Appendiks).



Figur 1

Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk.

Antal prøver: 718.

Antal samleværker: 114.

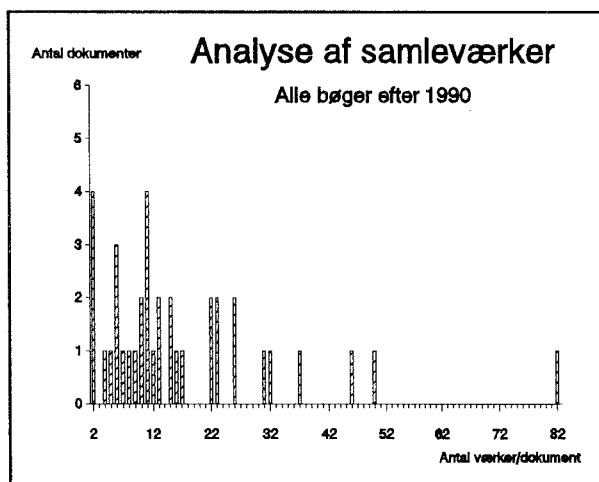
Stikprøve 2: Alle bøger publiceret efter 1990:

Man kunne forestille sig, at der har været en udvikling af publikations- eller indkøbspraksis gennem tiden. Dette ville da vise sig ved en stikprøveundersøgelse i litteraturen efter 1990 til primo 1992.

Der er 10.000 bøger på Danmarks Pædagogiske Bibliotek, der er færdigbehandlet fra denne periode.

Vi udtog første og sidste bind på hver hylde og udtog herved 211 bind, hvoraf 17% var samleværker. På dette grundlag kan vi med 95% sikkerhed forudsige at $17 \pm 5\%$ er samleværker af denne bogmasse.

Antallet af værker i samleværkerne er angivet på Figur 2. Det gennemsnitlige antal værker pr. samleværk er 17 ± 5 værker indenfor 95% grænsen.



Figur 2
Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk.
Antal prøver: 211
Antal samleværker: 36

Stikprøve 3: Engelsksprogede bøger publiceret 1980-1989:

Da mange kongresrapporter er på engelsk og disse rapporter har et betydeligt antal titler, udvalgte vi engelsk-sprogede bøger publiceret i en ti-års periode for at se, om dette gav anledning til signifikant variationer.

Dette udvalg blev foretaget maskinelt. Det drejede sig om ca. 16.000 bøger.

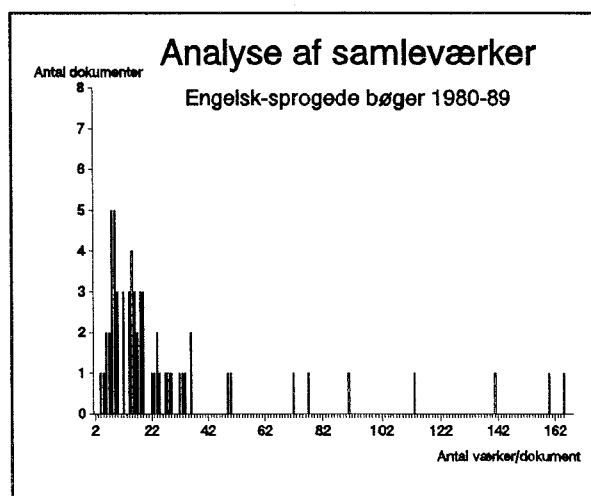
Stikprøven blev udvalgt ved at udtage poster med to tilfældig valgte to-cifrede endetal i ISBN nummeret.

Stikprøven viste, at 83 ud af 496 dokumenter var samleværker, hvilket giver 17% samleværker. 8 af disse var udlånt, da stikprøveundersøgelsen forgik, og der kunne ikke foretages en yderligere analyse af disse værker. Vi har derfor ikke medregnet disse, da forskellen er lille sammenlignet med den statistiske usikkerhed.

Regner vi med 75 samleværker, kan vi forudsige at $15 \pm 3\%$ af engelsk-sprogede publikationer

publiceret i 1980-89 med 95% sikkerhed er samleværker.

Samleværkernes bidrag til det totale antal værker er gengivet på Figur 3. Det gennemsnitlige antal værker pr. samleværk var 29 og vi kan med 95% sikkerhed sige, at ovennævnte del af samlingen indeholder gennemsnitlig 29 ± 9 værker pr. samleværk.



Figur 3
Stikprøveanalyse af andelen af samleværker samt antallet af værker i hvert samleværk.
Antal prøver: 496
Antal samleværker: 75

Sammenfatning af de tre undersøgelser.

Vi ser, at andelen af samleværker er meget konstant - knap 20% - for de tre stikprøver.

Antallet af titler pr. samleværk varierer som forventet, men variationen ligger indenfor de angivne usikkerheder. Gennemsnitlig er omkring 20 titler i et samleværk.

Det betyder, at Danmarks Pædagogiske Bibliotek i stedet for at præsentere ca. 8.000 nye titler om året kunne præsentere 40.000 nye titler om året uden at købe én bog mere.

Resursekravet til denne berigelse af online kataloget er $\frac{1}{2}\%$ af et værks procesomkostninger for hver ekstra titel (Poulsen 1992).

For Danmarks Pædagogiske Biblioteks vedkommende betyder det således, at omkostningerne ved en sådan berigelse af bibliotekets kataloger blot kræver en forøgelse på 2% af procesomkostningerne til monografierne.

Konklusion

Resultatet af disse stikprøveanalyser viser, at for hver bogtitel, der er registreret i Danmarks Pædagogiske Biblioteks online katalog, skjuler der sig 4 titler fra bidrag i samleværker.

En tilsvarende amerikansk undersøgelse af et college bibliotek viste at dér kunne antallet af titler 6-dobles. Det kan skyldes, at engelsk er det dominerende sprog i konferencelitteraturen, der bidrager væsentligt dels til antallet af samleværker og dels til det gennemsnitlige antal bidrag i disse samleværker. Denne effekt har ikke kunnet eftervises statistisk signifikant i de foretagne stikprøver på Danmarks Pædagogiske Bibliotek.

Omkostningerne ved at berige bibliotekets kataloger med 400% flere titler er lig med den arbejdstid, der anvendes til at behandle kun 2% af monografierne.

Tak

Til Frode Rasmussen for hjælp til at få statistikken genopfrisket samt kritik.

Litteraturliste

Arley, Niels & K.Rander Buch: Sandsynlighedsregning og statistik. København: Gjellerup. 1961.

Hoffman, Herbert H. and Jeruel L. Magner: Future Outlook: Better Retrieval through Analytic Catalogs. The Journal of Academic Librarianship, vol.11(1985), no.2, p.151-153.

Poulsen, Claus: On-line kataloger med indholdsfortegnelser af samleværker I: Kvalitetsforbedring for brugeren uden stort ressourceforbrug. Upubliceret. København, 1992.

APPENDIKS

Om statistikken

I det følgende opridses det statistiske grundlag for artiklens stikprøveanalyser. Som gennemgående eksempel anvendes artiklens måling på ca. 200.000 monografier, publiceret fra lige før århundredeskiftet til 1990. Bøgerne er systematisk opstillet, og ved at udtage sidste bind på 4' og 5' hylde i hver reol sikrer vi os, at stikprøven ikke udvælger bestemte dokumenttyper. Vi måler antallet af værker i hvert enkelt dokument i stikprøven, og stikprøven udtager 718 dokumenter.

Andelen af samleværker

Ud af de $N=718$ dokumenter var $N_s=114$ samleværker, det vil sige dokumenter der indeholder 2 eller flere værker.

Stikprøven viste sig altså at indeholde

$$x = \frac{N}{N_s} 100\% = 15,88\% \text{ samleværker.}$$

En fornyet stikprøve vil selvfølgelig med stor sandsynlighed have en anden andel af samleværker, men denne variation er begrænset af, at en vilkårlig ny prøve med 95% sandsynlighed vil ligge i et interval omkring den fundne værdi på det dobbelte af spredning på x , $2\sigma(x)$, der er på 1,36%.

Det vil sige at vi med 95% sikkerhed kan sige at antallet af samleværker i de ca. 200.000 bind ligger i intervallet $16 \pm 3\%$.

Årsagen til at vi kan udtale os om dette er skyldes to forhold. For det første er stikprøven binomialt fordelt, hvilket betyder at spredningen på hyppigheden x er

$$\sigma(x) = \sqrt{\frac{x(1-x)}{N}}$$

For det andet er såvel antallet af dokumenter xN og $(1-x)N$ i stikprøvens kategorier tilstrækkelig store til at fordelingen nærmer sig en normal fordeling, hvor netop 95% af stikprøverne vil ligge indenfor et interval på to gange spredningen omkring gennemsnittet fra den første stikprøve.

2. Antallet af værker i samleværkerne.

Ved hjælp af stikprøven kan vi også udtale os om antallet af bidrag fra de 114 samleværker. Antallet af bidrag tælles, $r(s)$, hvor s angiver at der er tale om et samleværk. Vi får en fordeling af et antal samleværker $N(s,r)$, hvor $r=2,3,4,5,\dots$ angiver antallet af bidrag fra hvert samleværk.

Hver af disse fordelinger er - som før - binomiale fordelinger. Fordelingerne er uafhængige af hinanden og har hver deres middelværdi og spredning. For at finde det gennemsnitlige antal bidrag pr. samleværk, middelværdien Y , summeres disse fordelinger og divideres med den samlede mængde af samleværker i stikprøven:

$$Y = \sum \frac{rN_s(r)}{N_s}, \text{ hvor } N_s = \sum_{r=2}^{\infty} N_s(r)$$

I den aktuelle stikprøve - se Figur 1 - er middelværdien $Y = 20,04$.

For de værdier af r , der er tale om her, gælder det - ifølge den såkaldte centrale grænseværdisætning - at denne sum er praktisk taget normalt fordelt. Og når middelværdien Y selv er praktisk taget normalt fordelt kan vi udlede, at middelværdien Y ved gentagelse af stikprøven med 95% sikkerhed vil ligge indenfor et interval på to gange spredningen på **middelværdien** omkring den første værdi af Y . Spredningen på middelværdien Y kan vises at være

$$\sigma(Y) = \sqrt{\frac{\sum_{r=2}^{\infty} (rN_s(r) - Y)^2}{N_s(N_s - 1)}}$$

- se Arley & Buch(1961) §11.7.

I den aktuelle stikprøve er $\sigma(Y) = 2,12$.

Det vil sige, at vi med 95% sikkerhed er mellem 16 og 24 værker pr samleværk.

Ovenstående betragtninger over statistikken er udlukkende eksemplificeret ved hjælp den ene af de tre udførte analyser, men gælder naturligvis for dem alle.