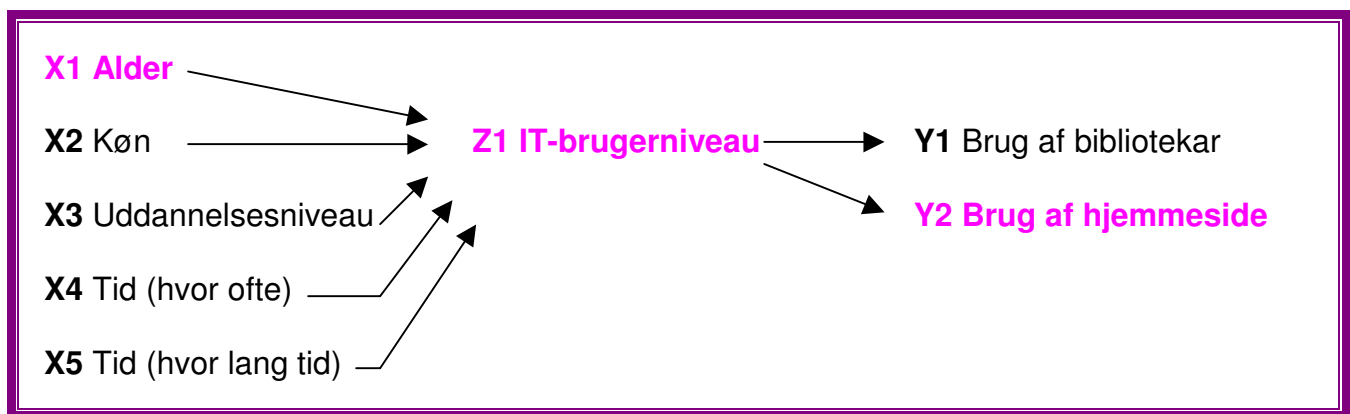


# Har teknologien overhalet bibliotekaren indenom?

## Samvariation?



## P4 Forskningsmetode

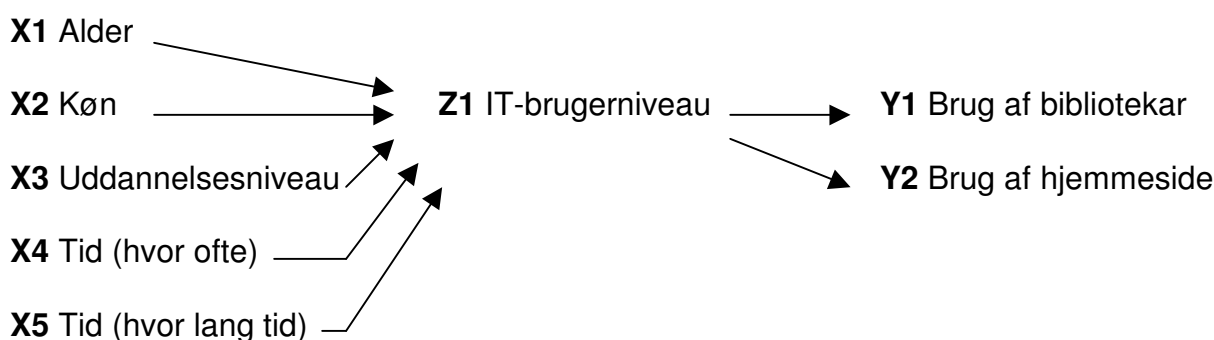
Af  
Anne-Marie Fiala Carlsen og Rikke Frisenette  
Hold 3, 2006  
Underviser: Kambiz K. Hormoozi  
Afleveret d. 29. november 2007  
Antal ord: 3525

## Indholdsfortegnelse

1 Indledning.....	3
2 Metode .....	3
3 Omkodning.....	4
4 Krydstabeller .....	8
5 Bivariat krydstabel.....	8
6 Procentdifferencer .....	11
7 Proportionsdifferencer .....	12
8 Chi-i-anden .....	13
9 Trivariat krydstabel.....	15
10 Goodman og Kruskal's tau .....	16
11 Procent- og proportionsdifference.....	18
12 Kausalanalyse.....	20
13 Konklusion .....	21
14 Litteraturhenvisninger.....	23
Bilag 1 .....	23
Bilag 2.....	24
Bilag 3.....	30

## 1 Indledning

Da vi i foråret læste Institut for Fremtidforskning's artikel med titlen "Zombier – de levende døde" og fandt bibliotekarfaget på dødslisten, blev vi inspireret til at undersøge om der er en sammenhæng mellem lånernes IT-brugerniveau og deres benyttelse af bibliotekaren og/eller bibliotekets hjemmeside. Resultatet blev en kausalmodel, figur 1 og en spørgeskemaundersøgelse, som vi gennemførte på Helsingør Hovedbibliotek i august 2007.



**Figur 1 Kausal model**

Kilde: P2

I P3 startede vi arbejdet med vores indsamlede data og beskrev to variabler. Nu ønsker vi at gå skridtet videre. Med udgangspunkt i vores kausalmodel fra P2, vil vi i denne opgave kigge på to, henholdsvis tre variabler, og undersøge om de samvarierer, altså om det er muligt statistisk at vise, at der er en sammenhæng mellem de valgte variabler<sup>1</sup>, således som vi har påstået i P2.

(fælles)

## 2 Metode

Vores kausalmodel består af flere uafhængige, én mellemliggende og to afhængige variabler. Vi har, for at holde os indenfor opgavens rammer, begrænset vores analyse til at omhandle den uafhængige variabel *alder*, den mellemliggende variabel *IT-brugerniveau* og den afhængige variabel *brug af bibliotekets hjemmeside*.

(fælles)

<sup>1</sup> Alstrup, Erik (2006) s. 200

### 3 Omkodning

Selv en lille spørgeskemaundersøgelse som vores resulterer i en ganske stor mængde tal og data. Men disse informationer er ikke meget værd, hvis de ikke fremtræder i en overskuelig, begribelig og relevant form<sup>2</sup>. Det kan derfor være nødvendigt at forenkle sine data. Allerede i P3 prøvede vi dette, da vi på baggrund af vores datamatrice, skabte frekvenstabeller og tabeller for målene central tendens, spredning, skævhed og kurtose. Et andet eksempel på dataforenkling er procentuering, som medfører en relativisering af absolutte værdier<sup>3</sup>. Procentuering kommer især til sin ret når relationen mellem to eller flere variabler skal beskrives<sup>4</sup>, og vi vil da også benytte os af metoden senere i vores opgave.

En anden metode, hvor der i mange tilfælde også er tale om dataforenkling, er omkodning. Helt overordnet dækker omkodning over enhver forandring af data, som indebærer at der dannes nye variable<sup>5</sup>. Desværre er omkodning ikke gratis og indebærer ofte også datareduktion<sup>6</sup>. Det er vigtigt at være opmærksom på, at når en variabel omkodes er måleniveauet ikke længere det samme<sup>7</sup>. I nogle tilfælde er det muligt ved hjælp af omkodning at danne variable, der ligger på et højere måleniveau end det oprindelige, men i de fleste tilfælde medfører det dog et lavere måleniveau<sup>8</sup>. Til gengæld er dataforenkling og omkodning et vigtigt skridt på vejen til at få svar på sine spørgsmål. Indeholder problemformuleringen spørgsmål, der involverer en analyse af sammenhængen mellem flere variable, kommer man ikke udenom dataforenkling og omkodning<sup>9</sup>. Omkodning kan ligeledes være nødvendigt for at opfylde krav til bestemte analysemodeller eller hvis talmaterialet er begrænset<sup>10</sup>. Ønskes en sammenligning af nye og gamle data, kan det ved hjælp af omkodning sikres at resultaterne kan sammenholdes<sup>11</sup>.

---

<sup>2</sup> Alstrup, Erik (2007a) s.3

<sup>3</sup> Alstrup, Erik (2007a) s.9

<sup>4</sup> Alstrup, Erik (2007a) s.9

<sup>5</sup> Alstrup, Erik (2006) s.111

<sup>6</sup> Alstrup, Erik (2006) s.118

<sup>7</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.49

<sup>8</sup> Alstrup, Erik (2006) s.123

<sup>9</sup> Alstrup, Erik (2007a) s.5

<sup>10</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.49

<sup>11</sup> Alstrup, Erik (2007a) s.5

Der findes seks teknikker til at omkodning af variabler<sup>12</sup>. De tre der muliggør sammenligning er

1. Logaritmisk transformation
2. Standardisering
3. Relativisering

Som hjælp til at forenkle, overskueliggøre og nå relevante svar benyttes

4. Kategorisering
  - a. Gruppering
  - b. Skalering
5. Indeksering
6. Dikotomisering

Vi har på baggrund af vores kausalmodel valgt at omkode variablerne *alder*, *IT-brugerniveau* og *brug af hjemmeside*. Vi arbejder med en uafhængig, en mellemliggende og en afhængig variabel. Det skal hjælpe os i vores videre arbejde med krydstabeller og senere kausalanalyse.

For alle tre variablers vedkommende består omkodningen i gruppering. Gruppering kan ske for variabler på alle måleniveauer<sup>13</sup>, og foretages ud fra to hovedideer. Enten, at der tages hensyn til en ligelig fordeling af værdier eller frekvenser, eller at fordelingen sker så det er hensigtsmæssigt i forhold til problemformuleringen<sup>14</sup>. I P3 fandt vi ud af, at vores frekvensfordeling for variabelen *alder* ikke er langt fra at være normalfordelt. Vi benytter derfor i første omgang kriteriet om ligelig fordeling og grupperer ved hjælp af standardafvigelsen. Hermed dannes tre grupper

1. Værdier som ligger under gennemsnittallet – standardafvigelsen:  $46-17 = 15-29$  år
2. Værdier som ligger mellem gennemsnittallet +/- standardafvigelsen:  
 $46\pm 17 = 30-63$  år
3. Værdier som ligger over gennemsnittallet + standardafvigelsen:  $46+17 = 64-81$  år<sup>15</sup>

---

<sup>12</sup> Alstrup, Erik (2006) s.110

<sup>13</sup> Alstrup, Erik (2006) s.164

<sup>14</sup> Alstrup, Erik (2006) s.113

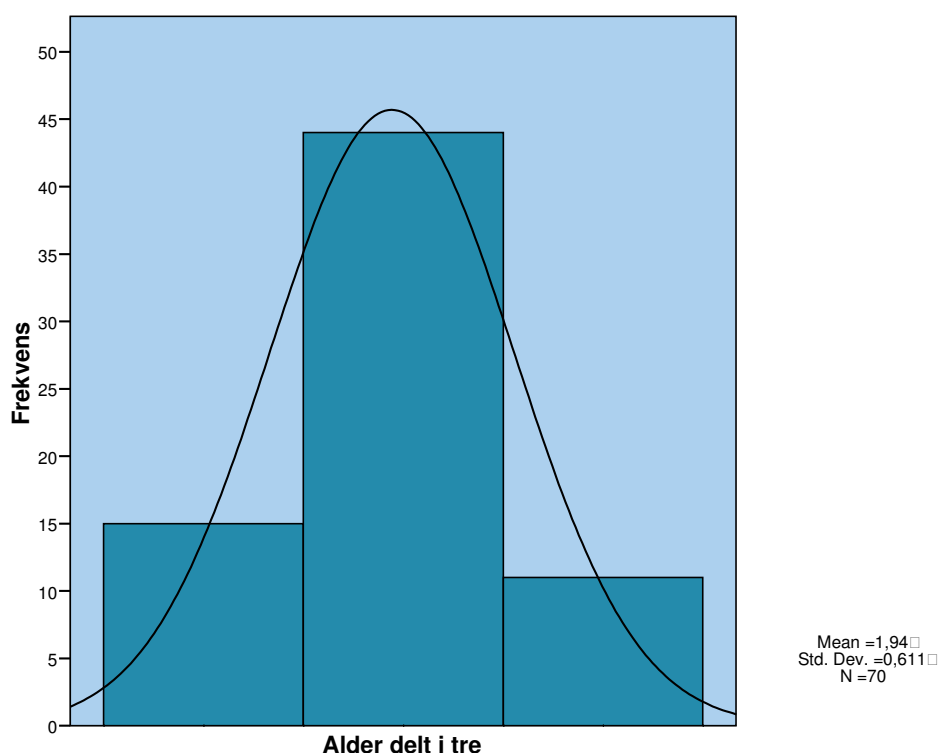
<sup>15</sup> Alstrup, Erik (2006) s.116

Dette resulterer i følgende frekvenstabel og histogram

**Tabel 3.1** Frekvenstabel over alder grupperet efter standardafvigelse

		Frekvens	Procent	Valid Procent
Valid	15-29 år	15	21,4	21,4
	30-63	44	62,9	62,9
	64-81	11	15,7	15,7
	Total	70	100,0	100,0

Kilde: Spørgeskemaundersøgelse august 2007



**Figur 2** Histogram

Kilde: Spørgeskemaundersøgelse august 2007

Vi kan se at normalfordelingskurven er klokkeformet og meget tæt på at være symmetrisk, men dog en anelse venstreskæv. En normalfordeling giver mulighed for mere avancerede analyser<sup>16</sup>, men vi er her mere interesserede i den eventuelle forbindelse der er mellem variablerne, og grupperer derfor på ny som det er relevant i forhold til vores spørgsmål.

<sup>16</sup> Alstrup, Erik (2006) s.117

Variablen *alder* vælger vi at dele i to, 15-59 år og 60-81 år, tabel 3.2. Dette er ikke en opdeling i yngre/ældre i aldersmæssig forstand. Vi er interesserede i sammenhængen mellem IT-brugerniveau og alder og har lavet gruppering ud fra en antagelse om, at man i dag skal være over 60 år for ikke at besidde et vist IT-brugerniveau. Dette underbygges af vores kassedigram fra P3<sup>17</sup>. Her fremgår det tydeligt, at majoriteten er respondenterne med "ingen" eller "lav" IT-erfaring skal findes i aldersgruppen 60+. Måleniveauet for variabelen alder er nu nominalt.

**Tabel 3.2 Frekvenstabel over alder grupperet**

		Frekvens	Procent	Valid Procent
Valid	15-59 år	55	78,6	78,6
	60-81 år	15	21,4	21,4
	Total	70	100,0	100,0

Kilde: Spørgeskemaundersøgelse august 2007

Respondenternes *IT-brugerniveau* har vi også omkodet. Her har vi også grupperet i to grupper, højt IT-niveau og lavt IT-niveau, Tabel 3.3. For at kunne tydeliggøre vores formodninger i krydstabellerne, har vi valgt at dele grupperne, så vi får respondenter med "meget erfaring" og "superbruger" i en gruppe for sig. Efter omkodningen er IT-brugerniveau en nominal variabel.

**Tabel 3.3 Frekvenstabel over IT-brugerniveau grupperet**

		Frekvens	Procent	Valid Procent
Valid	Højt IT niveau	32	45,7	45,7
	Lavt IT niveau	38	54,3	54,3
	Total	70	100,0	100,0

Kilde: Spørgeskemaundersøgelse august 2007

Til sidst har vi grupperet vores afhængige variabel *brug af bibliotekets hjemmeside*. Vi har foretaget en todeling, for ligesom med de andre variabler at forenkle og tydeliggøre, se Tabel 3.4. Hermed får vi en nominal variabel.

<sup>17</sup> Se P3, Figur 5 s. 16

**Tabel 3.4 Frekvenstabel over brug af hjemmeside grupperet**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Ofte	33	47,1	47,1	47,1
Sjældent	37	52,9	52,9	100,0
Total	70	100,0	100,0	

Kilde: Spørgeskemaundersøgelse august 2007

Omkodningen skal vi bruge, nå vi efterfølgende skal konstruerer krydstabeller.

(Rikke)

## 4 Krydstabeller

I P3 beskæftigede vi os med frekvenstabeller af de forskellige variabler. Krydstabeller er bi-, tri- eller multivariate frekvenstabeller<sup>18</sup>. Der er flere formål med krydstabeller, her vil vi beskæftige os med et af dem, nemlig "hjælp til at bedømme, om to eller flere variabler spiller mere eller mindre sammen"<sup>19</sup>. Vi arbejder videre med de tre omkodede variabler *alder*, *IT-brugerniveau* og *brug af hjemmesiden*.

I første omgang vælger vi den uafhængige og den afhængige variabel til en bivariat krydstabel.

(Anne-Marie)

## 5 Bivariat krydstabel

Krydstabeller kan behandles symmetrisk og asymmetrisk<sup>20</sup>. Hvis vi bare vil se fordelingen af to variabler og ikke beskæftige os med, om den ene variabel påvirker den anden, vil det være symmetrisk behandling af variablerne<sup>21</sup>, men da vi netop tager udgangspunkt i at undersøge samspillet mellem vores variabler – påvirker den ene variabel den anden, men ikke omvendt<sup>22</sup> - er det en asymmetrisk sammenhæng mellem variablerne. Ud fra vores kausalmodel, har vi valgt at lave en krydstabel over *alder* og *brug af hjemmeside*, hvor

<sup>18</sup> Alstrup, Erik (2006) s. 176

<sup>19</sup> Alstrup, Erik (2006) s. 176

<sup>20</sup> Alstrup, Erik (2006) s. 177

<sup>21</sup> Alstrup, Erik (2006) s. 176

<sup>22</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s. 83

*alder* er den uafhængige variabel og *brug af hjemmeside* er den afhængige. Ens alder kan påvirke ens brug af bibliotekets hjemmeside, men ikke omvendt.

Kausalmodellen ser herefter ud som i figur 3:



**Figur 3** Kausalmodel med en uafhængig variabel og en afhængig variabel

Laver vi en krydstabel med de oprindelige variabler, får vi en stor og uoverskuelig tabel. Eftersom vi har respondenter i alderen fra 15 til 81 år og fem svarmuligheder i brug af hjemmesiden vil vi få mange tomme celler og vil have svært ved at bruge tallene til noget. Samtidig hermed skal vi huske at tage hensyn til, at vores datagrundlag er lille, kun 70 respondenter. Det betyder, at hvis vi skal kunne måle om vores variabler samvarierer, er der nogle tommelfingerregler for, hvor store cellefrekvenserne bør være. Der er dog forskellige meninger om, hvor stort antallet af frekvenser skal være, nogen siger 20, andre 15 og andre igen 5<sup>23</sup>. Vi accepterer frekvensværdier ned til 15.

Vi har derfor omkodet både respondenternes *alder* og *brug af hjemmesiden* og kan nu lave vores første krydstabel.

Først fremkommer en tabel (Tabel 5.1), der viser hvor mange af vores respondenter der har svaret på begge spørgsmål. I vores tilfælde har alle svaret.

**Tabel 5.1**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Hvor ofte benyttes bibliotekets hjemmeside grupperet * Alder	70	100,0%	0	,0%	70	100,0%

Kilde: Spørgeskemaundersøgelse august 2007

<sup>23</sup> Alstrup, Erik (2006) s. 182

Herefter kommer selve krydstabellen (Tabel 5.2), hvor vi direkte kan aflæse, at ud af de respondenter der ofte benytter hjemmesiden er 25 mellem 15-59 år. Ud af 70 respondenter er der i alt 33 der ofte benytter hjemmesiden og 37 der sjældent benytter den. Vi kan også se, at der er 55 respondenter i aldersgruppen 15-59 år og 15 i gruppen 60-81 år. Det er altså de faktuelle tal vi her ser.

**Tabel 5.2 Hvor ofte benyttes bibliotekets hjemmeside \* Alder**

**Krydstabel**

		Alder		Total
		15-59 år	60-81år	
Hvor ofte benyttes bibliotekets hjemmeside grupperet	Ofte	25	8	33
	Sjældent	30	7	37
Total		55	15	70

Kilde: Spørgeskemaundersøgelse august 2007

For bedre at kunne sammenligne data anvender vi den dataforenklingsteknik der hedder procentuering<sup>24</sup>. Vi beder altså SPSS om at udregne procenter for de enkelte tal i krydstabellen.

Der er flere måder at gøre dette på. Vælger vi som her, at den uafhængige variabel er placeret i kolonnen, skal vi også procentuere kolonnevis, eller vertikalt. Det betyder at hver kolonne tilsammen bliver 100%. Procentueringsretningen følger det valg vi tager for placering af den uafhængige variabel<sup>25</sup>.

I SPSS er der også mulighed for at procentuere horisontalt eller totalt, vi kommer ikke nærmere ind på disse muligheder her.

<sup>24</sup> Alstrup, Erik (2006) s. 168-169

<sup>25</sup> Alstrup, Erik (2006) s. 178

**Tabel 5.3 Hvor ofte benyttes bibliotekets hjemmeside \* Alder****Procentueret****Krydstabel**

			Alder		Total
			15-59 år	60-81 år	
Hvor ofte benyttes bibliotekets hjemmeside grupperet	Ofte	Count	25	8	33
		% within Alder	45,5%	53,3%	47,1%
	Sjældent	Count	30	7	37
		% within Alder	54,5%	46,7%	52,9%
Total		Count	55	15	70
		% within Alder	100,0%	100,0%	100,0%

Kilde: Spørgeskemaundersøgelse august 2007

I Tabel 5.3 ser vi nu på procenterne, som giver et bedre overblik, end de absolutte tal fra Tabel 5.2, og vi kan nu udregne procentdifferencerne, og siden proportionsdifferencerne, i vores krydstabel.

*(Anne-Marie)*

## 6 Procentdifferencer

For at kunne se nærmere på, om vores variabler samvarierer, kan vi benytte procentdifferencer<sup>26</sup>

Procentdifferencen viser forskellen mellem to procentuerede tal i krydstabellen<sup>27</sup>. I vores tilfælde har vi valgt vertikal procentuering og derfor skal tabellen læses horisontalt<sup>28</sup>.

Procentdifferencen er forskellen mellem de to procenter i hver horisontal række.

Vi har lavet en 2x2 tabel, en såkaldt firefeltstabel, der kun omfatter to rækker og to kolonner. Dette medfører, at vores procentdifferencer bliver af samme størrelsesorden, men med modsat fortegn. Procentdifferencen beskrives da med betegnelsen epsilon, der beskriver netop samme procentdifferens med forskelligt fortegn<sup>29</sup>. I Tabel 5.3 er procentdifferensen epsilon 7,8, altså henholdsvis -7,8 og +7,8. Fortegnene fortæller i hvilken retning sandsynligheden peger; ud af de 47,1% af respondenterne, der ofte bruger hjemmesiden, er det mere sandsynligt at en tilfældig udtrukket respondent er ældre (60-81 år) end yngre (15-59 år). Det kan være svært at sige noget generelt om, hvornår en

<sup>26</sup> Alstrup, Erik (2007) s. 169

<sup>27</sup> Alstrup, Erik (2006) s. 201

<sup>28</sup> Alstrup, Erik (2006) s. 180

<sup>29</sup> Alstrup, Erik (2006) s. 187 og 201

procentdifference er stor eller lille, det afhænger af hvor stort datagrundlaget er. Da vores er et lille grundlag, skal der ikke meget til at flytte retningen. Havde blot én respondent i gruppen 60-81 år svaret "sjældent" ville der stort set ikke være nogen procentdifference<sup>30</sup>.

(Anne-Marie)

## 7 Proportionsdifferencer

Proportionsdifferencer er de procentuerede frekvenstal divideret med 100, og udtrykkes i en proportion med tre decimaler<sup>31</sup>. De har interesse, fordi vi her kan aflæse styrken "hvormed en given egenskab påvirker ændringen i sandsynligheden for, at en anden egenskab indtræffer"<sup>32</sup>.

Proportionsdifferencen i Tabel 5.3 udregnes således:

Cellefrekvens 25 : marginalfordelingen 55 respondenter = 0,455

Cellefrekvens 8 : marginalfordelingen 15 respondenter = 0,533

Proportionsdifferencen er  $0,455 - 0,533 = -0,078$

Da det er en firefeltstabel, ved vi at de to differencer er ens blot med modsat fortegn. Tabel 7.1 viser proportionerne og proportionsdifferencerne på Tabel 5.3:

**Tabel 7.1 Proportioner og proportionsdifferencer**

	Proportioner		Proportionsdifferencer
	15-59 år	60-81 år	
Ofte	0,455	0,533	-0,078
Sjældent	0,545	0,467	0,078

Kilde: Tabel 5.3

Procent- og proportionsdifferencer kan kun vise om der er samvariation ikke hvor stærk den er<sup>33</sup>. Det samme gør sig gældende for det statistiske mål for samvariation, chi-i-anden.<sup>34</sup>

(Anne-Marie)

<sup>30</sup> Pors, Niels Ole (1993) s. 23

<sup>31</sup> Alstrup, Erik (2006) s. 201

<sup>32</sup> Alstrup, Erik (2006) s. 189-190

<sup>33</sup> Alstrup, Erik (2006) s. 190

<sup>34</sup> Alstrup, Erik (2006) s. 242

## 8 Chi-i-anden

Chi-i-anden er en statistisk test, der bruges til at vurdere den statistiske signifikans af et resultat. Det vil sige, at man bruger testen til at undersøge, om resultaterne i undersøgelsen forekommer oftere end det kan forventes ud fra en tilfældighed<sup>35</sup>.

Signifikanstest er bygget op omkring begrebetsparret 0-hypotesen og arbejdshypotesen<sup>36</sup>. 0-hypotesen bygger på, at en given variabel ikke har nogen sammenhæng med en anden variabel. Arbejdshypotesen bygger derimod oftest på en formodning om sammenhæng<sup>37</sup>. Signifikanstesten udmunder i to talværdier, teststørrelsen og signifikanssandsynligheden<sup>38</sup>. Teststørrelsen konstrueres som forskellen mellem undersøgelsens faktiske tal og en forventet fordeling, som er 0-hypotesen. Teststørrelsen bruges ikke i forbindelse med chi-i-anden. Det gør derimod den tilknyttede signifikanssandsynlighed, som benævnes p-værdien<sup>39</sup>. P-værdien er et mål for sandsynligheden af den faktiske frekvensfordeling, hvis vi forudsætter at 0-hypotesen er sand<sup>40</sup>. Eller med andre ord, et mål for sandsynligheden af, at resultaterne kunne være opstået tilfældigt. P-værdien udtrykkes oftest procentuelt<sup>41</sup> og skillepunktet vurderes af de fleste eksperter til at gå ved 5%<sup>42</sup>. Da det jo er 0-hypotesen der testes, vil det sige at en p-værdi mindre end 5% ( $< 0,05$ ) betegnes som signifikant og at arbejdshypotesen kan accepteres. Derimod betyder en p-værdi højere end 5% en forkastelse af arbejdshypotesen<sup>43</sup>.

Chi-i-anden kan benyttes på alle måleniveauer, men er mest hensigtsmæssig på nominalniveau, da der findes andre og stærkere mål for samvariation mellem variabler på ordinal- og skalaniveau<sup>44</sup>.

---

<sup>35</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.65

<sup>36</sup> Alstrup, Erik (2006) s.203

<sup>37</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.65

<sup>38</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.66

<sup>39</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.91

<sup>40</sup> Alstrup, Erik (2006) s.203

<sup>41</sup> Alstrup, Erik (2006) s.192

<sup>42</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.66

<sup>43</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.66

<sup>44</sup> Oversigt 18

Tabel 8.1 viser chi-i-anden for vores krydstabel med to variabler.

**Tabel 8.1 Chi-i-anden**

	Værdi	df	Asymtotisk signifikans (tosidet)
Pearson Chi-Square	,294(a)	1	,588
Antal valide undersøgelsesenheder	70		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 7,07.

Kilde: Spørgeskemaundersøgelse august 2007

Tabellen indeholder tre talværdier. *Værdi* dækker over selve den beregnede værdi for chi-i-anden<sup>45</sup>. Fordi chi-i-anden-værdien beregnes ud fra henholdsvis den observerede frekvensfordeling og den forventede frekvensfordeling, altså 0-hypotesen, kan man sige at jo større chi-i-anden er, jo større er forskellen mellem de to frekvensfordelinger, og jo større er sandsynligheden for samvariation<sup>46</sup>. Vores resultater viser en lav værdi på 0,294, hvilket ikke indikerer samvariation mellem de to variabler. Dette underbygges af endnu et tal, nemlig den *asymtotiske signifikans (tosidet)*, som dækker over p-værdien. Vi har en p-værdi på 58,8%. Dette er temmelig langt fra signifikansgrænsen på 5%, og sandsynligheden for, at der er hold i 0-hypotesen er altså temmelig stor. Blot ved at se på p-værdien, kan vi konkludere, at der ikke er samvariation.

At der her er tale om en to-sidet p-værdi vil sige, at der ikke er taget hensyn til om en eventuel samvariation er positiv eller negativ<sup>47</sup>. Kolonnen med overskriften *df* dækker over *degrees of freedom/frihedsgrader*<sup>48</sup>. Vores krydstabel er lavet på baggrund af én uafhængig og én afhængig variabel, og derfor er df-værdien en<sup>49</sup>. Det sidste tal 70, er antallet af respondenter.

Vi inddrager nu vores mellemliggende variabel.

(Rikke)

<sup>45</sup> Alstrup, Erik (2006) s.239

<sup>46</sup> Alstrup, Erik (2006) s.240

<sup>47</sup> Alstrup, Erik (2006) s.240

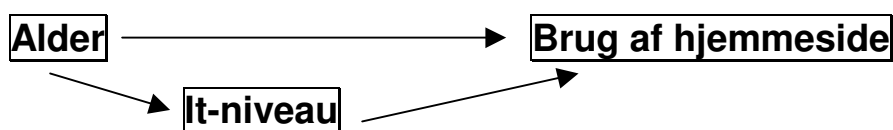
<sup>48</sup> Alstrup, Erik (2006) s.239

<sup>49</sup> Alstrup, Erik (2006) s.240

## 9 Trivariat krydstabel

Vi konstruerer en krydstabel med tre variabler. Vores kausalmodel indeholder en mellemliggende variabel, *IT-brugerniveau*. Ved at danne en trivariat frekvenstabel, kan denne variabel bruges som kontrolvariabel og vi kan undersøge om der er direkte eller indirekte samvariation.

Kausalmodellen ser herefter ud som i Figur 4:



**Figur 4** Kausalmodel med en uafhængig, en mellemliggende og en afhængig variabel

Vi bruger den omkodede variabel *IT-brugerniveau* som den konstante variabel.

Vi tror alder kan have en indirekte effekt på brugen af hjemmesiden, formidlet gennem IT-brugerniveauet<sup>50</sup>.

Vores krydstabel med de tre variabler (Tabel 9.1) viser bl.a. at respondenter med lavt IT-brugerniveau sjældent bruger hjemmesiden, næsten 66% uanset alder. Tilsvarende viser tabellen, at 62,5% af respondenterne med højt IT-brugerniveau ofte benytter hjemmesiden, uanset alder. Kigger vi på alderen, ser vi f.eks. at 62,1% af de 15-59 årige ofte benytter hjemmesiden og har et højt IT-brugerniveau. 50% af aldersgruppen 60-81 der ofte benytter hjemmesiden, har et lavt IT-brugerniveau og 73,1% med lavt IT-brugerniveau som sjældent bruger hjemmesiden er mellem 15-59 år. Vi kan altså aflæse en del oplysninger direkte ud af tabellen og allerede her gætte på, at det ikke er variabelen alder der er bestemmende for en eventuel sammenhæng. I øvrigt det samme som vores chi-i-anden test viste.

<sup>50</sup> Alstrup, Erik (2006) s. 246

**Tabel 9.1 Hvor ofte benyttes bibliotekets hjemmeside \* Alder \* IT-bruger niveau****Krydstabel**

IT-bruger niveau grupperet				Alder		Total
				15-59 år	60-81 år	
Højt IT niveau	Hvor ofte benyttes bibliotekets hjemmeside grupperet	Ofte	Count	18	2	20
			% within Alder	62,1%	66,7%	62,5%
	Sjældent	Count	11	1	12	
		% within Alder	37,9%	33,3%	37,5%	
	Total		Count	29	3	32
			% within Alder	100,0%	100,0%	100,0%
Lavt IT niveau	Hvor ofte benyttes bibliotekets hjemmeside grupperet	Ofte	Count	7	6	13
			% within Alder	26,9%	50,0%	34,2%
	Sjældent	Count	19	6	25	
		% within Alder	73,1%	50,0%	65,8%	
	Total		Count	26	12	38
			% within Alder	100,0%	100,0%	100,0%

Kilde: Spørgeskemaundersøgelse august 2007

Nu vil vi undersøge hvilken betydning den mellemliggende variabel *IT-niveau* har for vores samvariation

(Anne-Marie)

## 10 Goodman og Kruskal's tau

Der findes ud over signifikanstesten chi-i-anden, andre måder at teste for samvarians. En mulighed kunne være at benytte associationsmålet lambda. Lambda har dog den ulempe, at målet er følsomt overfor ujævne eller skæve fordelinger, som dermed resulterer i lave lambdakoefficienter<sup>51</sup>. Netop dette viser vores beregninger, se Tabel 10.1, og benyttelsen af lambda er derfor uhensigtsmæssig. Vi vælger i stedet at tage udgangspunkt i

<sup>51</sup> Alstrup, Erik (2007b) Oversigt 18

associationsmålet Goodman og Kruskal's tau, som er konstrueret med henblik på nominalvariabler<sup>52</sup>.

**Tabel 10.1 Lambda og tau**

It-bruger niveau grupperet				Value	Asymp. Std. Error <sup>a</sup>	Approx. Sig.
Højt IT niveau	Nominal by Nominal	Lambda	Symmetric	,000	,000	. <sup>c</sup>
			Hvor ofte benyttes bibliotekets hjemmeside grupperet Dependent	,000	,000	. <sup>c</sup>
			Alder Dependent	,000	,000	. <sup>c</sup>
	Goodman and Kruskal tau	Hvor ofte benyttes bibliotekets hjemmeside grupperet Dependent	,001	,010	,878 <sup>d</sup>	
		Alder Dependent	,001	,010	,878 <sup>d</sup>	
Lavt IT niveau	Nominal by Nominal	Lambda	Symmetric	,000	,139	1,000
			Hvor ofte benyttes bibliotekets hjemmeside grupperet Dependent	,000	,266	1,000
			Alder Dependent	,000	,000	. <sup>c</sup>
	Goodman and Kruskal tau	Hvor ofte benyttes bibliotekets hjemmeside grupperet Dependent	,051	,074	,169 <sup>d</sup>	
		Alder Dependent	,051	,074	,169 <sup>d</sup>	

a. Not assuming the null hypothesis.

c. Cannot be computed because the asymptotic standard error equals zero.

d. Based on chi-square approximation

Kilde: Spørgeskemaundersøgelse august 2007

En fordel ved tau, som vi kan drage nytte af i vores undersøgelse er, at dette associationsmål ikke er følsomt overfor små datamængder<sup>53</sup> og vi har jo som bekendt kun 70 respondenter. Beregningen af tau bygger på en asymmetrisk tilgang, altså at der opereres med en afhængig og en uafhængig variabel<sup>54</sup>. Tau er et normeret mål og koefficienterne falder mellem 0 og 1, hvor 0 er udtryk for statistisk uafhængighed og 1 for komplet samvariation<sup>55</sup>. Hvad viser vores beregninger os så? I Tabel 10.1 ser vi at tau er beregnet for kontrolvariablens to værdier, lavt og højt IT-brugerniveau. I den yderste højre kolonne *approx. Sig.*, kan p-værdien aflæses. P-værdien for højt IT-brugerniveau er 87,8% og 16,9% for lavt IT-brugerniveau. Vi husker fra tidligere at p-værdien skal være

<sup>52</sup> Alstrup, Erik (2006) s.214

<sup>53</sup> Alstrup, Erik (2007b) Oversigt 18

<sup>54</sup> Alstrup, Erik (2006) s. 176

<sup>55</sup> Alstrup, Erik (2006) s.214

under 5% for at vores arbejdshypotese om sammenhæng kan accepteres<sup>56</sup>. Der er vi langt fra og vi kan derfor se bort fra de øvrige værdier som er udregnet i tabellen.

Vi kan også teste ved igen at udregne procent- og proportionsdifferencer.

(Rikke)

## 11 Procent- og proportionsdifference

Vi udregner procent – og proportionsdifferencer, denne gang både på øverste henholdsvis nederste deltabel og mellem de to deltabeller, se Tabel 11.1 og 11.2.

Vi kan tolke på de fremhævede resultater, at sandsynligheden for at træffe en 15-59 årig frem for en 60-81 årig, med højt IT-brugerniveau der ofte bruge hjemmesiden er reduceret med 4,6 procentpoint. Sandsynligheden for at træffe en 15-59 årig som ofte bruger hjemmesiden er øget med 35,2 procentpoint, hvis vedkommende har et højt IT-brugerniveau frem for lavt IT-brugerniveau<sup>57</sup>.

**Tabel 11.1 Procent- og proportionsdifferencer af Tabel 9.1 Horisontalt**

		Procentdifference	Proportionsdifference
Højt IT-niveau	Ofte	<b>-4,6</b>	-0,046
	Sjældent	4,6	0,046
Lavt IT-niveau	Ofte	-23,1	-0,231
	Sjældent	23,1	0,231

Kilde: Tabel 9.1

**Tabel 11.2 Procent- og proportionsdifferencer af Tabel 9.1 Vertikalt**

		Procentdifference	Proportionsdifference
15-59 år	Ofte	<b>35,2</b>	0,352
	Sjældent	-35,2	-0,352
60-81 år	Ofte	16,7	0,167
	Sjældent	-16,7	-0,167

Kilde: Tabel 9.1

<sup>56</sup> Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004) s.66

<sup>57</sup> Alstrup, Erik (2006) s. 212

Vi ser på, om der er nogen effekt af graden af respondenternes IT-brugerniveau og laver to tabeller med procentandele, ud fra de tal vi fandt i Tabel 11.1 og 11.2, samt tallene fra Tabel 5.3. Hvis kontrolvariablen har betydning for, hvordan frekvenserne på den afhængige variabel fordeler sig, vil procentandelene, der er fremhævet til højre i Tabel 11.3 og Tabel 11.4, være mere eller mindre forskellig fra procentandelene til venstre i cellerne<sup>58</sup>.

**Tabel 11.3 Procentandele med Ofte brug af hjemmesiden fordelt på variablerne IT-brugerniveau og Alder (fra Tabel 9.1) sammenholdt med procentandelene ved krydstabuleringen af varianterne Alder og Brug af hjemmesiden (fra Tabel 5.3)**

	15-59 år			60-81 år		
I alt 70 respondenter	45,5%	Højt IT	<b>62,1%</b>	53,3%	Højt IT	<b>66,7%</b>
	45,5%	Lavt IT	<b>26,9%</b>	53,3%	Lavt IT	<b>50,0%</b>

Kilde: Tabel 5.3 og Tabel 9.1

**Tabel 11.4 Procentandele med Sjældent brug af hjemmesiden fordelt på variablerne IT-brugerniveau og Alder (fra Tabel 9.1) sammenholdt med procentandelene ved krydstabuleringen af varianterne Alder og Brug af hjemmesiden (fra Tabel 5.3)**

	15-59 år			60-81 år		
I alt 70 respondenter	54,5%	Højt IT	<b>37,9%</b>	46,7%	Højt IT	<b>33,3%</b>
	54,5%	Lavt IT	<b>73,1%</b>	46,7%	Lavt IT	<b>50,0%</b>

Kilde: Tabel 5.3 og Tabel 9.1

Ved alle disse udregninger er vi klar over, at flere af cellerværdierne i Tabel 9.1 er meget lave, hvilket øger vores usikkerhed.

Til sidst prøver vi at samle de forskellige beregninger i en kausalanalyse i tabelform.

*(Anne-Marie)*

<sup>58</sup> Alstrup, Erik (2006) s. 212

## 12 Kausalanalyse

Vi forestillede os i P2, at IT-brugerniveau kunne påvirke respondenternes brug af bibliotekets hjemmeside, men at der også er variabler der kan påvirke IT-brugerniveauet, f.eks. alder, køn og uddannelse. Derved blev IT-brugerniveau en mellemliggende variabel, og alder den variabel der kunne tænkes at have en indirekte effekt på brug af hjemmesiden via IT-brugerniveauet<sup>59</sup>.

Vi har prøvet at samle vore beregninger i en kausalanalyse i tabelform, Tabel 12.1.

**Tabel 12.1 Kausalanalyse i tabelform**

	Alder <sup>60</sup>	IT-brugerniveau <sup>61</sup>
Bivariat samvariation	7,8	28,3 <sup>62</sup>
Samlede kausaleffekter	7,8 og 7,8	35,2 og 16,7
Direkte kausaleffekter	4,6 og 23,1	35,2 og 16,7
Indirekte kausaleffekter	3,2 og -15,3	0
Spuriøse effekter	0	-6,9 og 11,6

Kilde: Tabel 5.3, Tabel 11.1, Tabel 11.2 og bilag 1

Den bivariate sammenhæng, den samlede kausaleffekt og den direkte effekt finder vi ud fra de tabeller vi har lavet i forvejen. Den indirekte og den spuriøse effekt finder vi ved hjælp af følgende udregninger:

$$\text{Indirekte kausaleffekt} = \text{samlet kausaleffekt} - \text{direkte kausaleffekt}$$

$$\text{Spuriøs effekt} = \text{bivariat sammenhæng} - \text{samlet kausaleffekt}^{63}$$

Kigger vi på tallene i Tabel 12.1, ser vi, at den bivariate samvariation for alder og brug af hjemmeside er lav (7,8), mens den for IT-brugerniveau og brug af hjemmeside er moderat (28,3)<sup>64</sup>. Den direkte effekt af IT-brugerniveauets påvirkning af de 15-59 årige er lidt højere endnu (35,2). Den direkte effekt af alderens påvirkning af respondenter med højt IT-

<sup>59</sup> Hellevik, Ottar (2003) s. 291

<sup>60</sup> Tallene til venstre i kolonnen med alder er for respondenter med højt IT-bruger niveau og til højre for dem med lavt IT-brugerniveau

<sup>61</sup> Tallene til venstre i kolonnen med IT-brugerniveau er de 15-59 årige og til højre de 60-81 årige

<sup>62</sup> For at kunne finde tallet for bivariat samvariation for IT-brugerniveau har vi fremstillet endnu en bivariat krydstabel, hvor IT-brugerniveau er den uafhængige variabel og Brug af hjemmesiden den afhængige. Se Bilag 1

<sup>63</sup> Hellevik, Ottar (2003) s. 293

<sup>64</sup> Alstrup, Erik (2006) s. 254

brugerniveau er lav (4,6) og af respondenter med lavt IT-brugerniveau er effekten moderat (23,1).

Med kausalanalysen er der mange flere muligheder, end vi kommer ind på her. Vi kunne have undersøgt for andre variabelers samvariation med vores uafhængige variabel, og derved kontrolleret hvor stort eller småt samspillet mellem variablerne i vores kausalmodel fra P2 er.

(Anne-Marie)

### 13 Konklusion

Med udgangspunkt i vores kausalmodel har vi nu analyseret tre af vores variabler. Da vi startede på P2, troede vi, at der var en sammenhæng mellem den uafhængige variabel *alder* og den afhængige variabel *brug af hjemmesiden*. Vi antog også at den mellemliggende variabel *IT-brugerniveau* ville have en betydning.

Analysen af den uafhængige og den afhængige variabel, vores bivariate krydstabel, viste, at der næsten ingen procentdifference var og at retningen faktisk var modsat det vi troede. Vi tjekkede med signifikans-testen chi-i-anden og også her så vi, at p-værdien var meget langt fra de 5% - der kunne ikke vises samvariation.

Så prøvede vi med vores mellemliggende variabel som kontrolvariabel. Ville det gøre nogen forskel? Den trivariate krydstabel gav nogle overraskende tal. Faktisk viste den, at respondenter med højt IT-brugerniveau, uanset alder, ofte benyttede hjemmesiden – næsten 2/3. Alder viste sig igen tydeligt, bare ikke som vi havde troet. Vi testede med både procent- og proportionsdifferencer og associationsmålet tau, samt satte tallene op i en kausalanalyse i tabelform, men vores tal var kun lave og moderate. Der er altså ingen samvariation.

For at lave kausalanalysen var vi nødt til at lave en krydstabel med IT-brugerniveau som uafhængig og brug af hjemmeside som afhængig variabel (Bilag 1). Denne viser lidt af vores antagelser, at respondenter med højt IT-brugerniveau oftere benytter hjemmesiden end respondenter med lavt IT-brugerniveau. Vi kunne være gået videre og have lavet en chi-i-anden test. Formodentlig kunne vi vise en samvariation her.

Også andre af vores variabler havde fortjent opmærksomhed, men der er desværre ikke plads til det, i denne opgave.

Et par fejlkilder vi har været opmærksomme på:

*IT-brugerniveau* grupperede vi ved "meget erfaring" og "superbruger". Havde vi taget gruppen med "middel erfaring" med, havde vores tal set helt anderledes ud. Det samme er tilfældet i vores deling af *brug af hjemmeside*, hvor vi ikke tog hensyn til, at gruppen af respondenter der aldrig bruger hjemmesiden, måske ikke bare kan sættes sammen med dem, der bruger den sjældent.

Allervigtigst er dog, at vores datagrundlag er meget lille og at vi derfor ikke kunne forvente at finde den store samvariation. Denne opgave er en øvelse i forskningsmetode og undersøgelsen og vores resultater kan derfor ikke betegnes som værende repræsentative.

*(Fælles)*

## 14 Litteraturhenvisninger

### Bøger:

Alstrup, Erik (2006). *Kan det være anderledes...?* – Metoder til at beskrive og analysere data. Danmarks Biblioteksskole

Hellevik, Ottar (2003). *Forskningsmetode i sosiologi og statsvitenskap*. Universitetsforlaget

Pors, Niels Ole (1993) *Statistiske metoder*. Introduktion for bibliotekarer. Danmarks Biblioteksskole

Ørum Andersen, Lars og Ull Jakobsen, Karsten (2004). *Introduktion til statistik med SPSS*. Samfundslitteratur

### Materiale udleveret i undervisningen:

Alstrup, Erik (2007a). *Dataforenkling og omkodningsteknikker i SPSS*

Alstrup, Erik (2007a). Oversigt 18

## Bilag 1

Hvor ofte benyttes bibliotekets hjemmeside grupperet \* It-bruger niveau grupperet Crosstabulation

			It-bruger niveau grupperet		Total
			Højt IT niveau	Lavt IT niveau	
Hvor ofte benyttes bibliotekets hjemmeside grupperet	Ofte	Count % within It-bruger niveau grupperet	20 62,5%	13 34,2%	33 47,1%
	Sjældent	Count % within It-bruger niveau grupperet	12 37,5%	25 65,8%	37 52,9%
Total		Count % within It-bruger niveau grupperet	32 100,0%	38 100,0%	70 100,0%

## Bilag 2

# Brug af bibliotekaren og bibliotekets hjemmeside

En spørgeskemaundersøgelse i Forskningsmetode

2. Semester på Danmarks Biblioteksskole

27.august 2007

Vi er to biblioteksstuderende fra Danmarks Biblioteksskole, der laver eksamensopgave om voksnes brug af bibliotekaren og bibliotekets hjemmeside.

Vi spørger til lidt forskelligt, om dig og din brug af nogle af bibliotekets muligheder.

Du kan spørge os, hvis du har forståelsesspørgsmål.

Din besvarelse er helt anonym, og vil kun blive brugt til denne undersøgelse.

**Med venlig hilsen Anne-Marie Fiala og Rikke Frisenette**

**Lidt om dig:**

1. Hvor gammel er du? \_\_\_\_\_ år

2. Køn Mand  1 Kvinde  2

3. Sidste afsluttede uddannelse Folkeskolen  1

(sæt kun ét kryds) Gymnasial uddannelse  2

Erhvervsuddannelse  3

Kort uddannelse  4

Mellemlang videregående uddannelse  5

Lang videregående uddannelse  6

**Lidt om dit niveau som IT-bruger:**

4. På en skala fra 0 til 5, hvor 0 er ”ingen erfaring” og 5 er ”superbruger”, hvordan vil du selv betegne dit IT-brugerniveau?

Sæt ring om det tal du synes passer bedst:

<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Ingen erfaring			Super bruger		

5. Har du daglig adgang til Internettet? Ja  1 Nej  2

**Din brug af biblioteket:**

6. Hvor ofte kommer du på biblioteket
- 1 eller flere gange om ugen  1
- 1-3 gange om måneden  2
- 1-5 gange på et halvt år  3
- 1-2 gange om året.  4
- 
7. Hvor lang tid bruger du oftest pr. besøg
- 0 - ½ time  1
- ½ - 1 time  2
- 1 - 1½ time  3
- 1½- 2 timer  4
- Mere end 2 timer  5
- 
8. I hvilken forbindelse bruger du biblioteket?
- (Sæt kryds ved de vigtigste, max. tre krydser)
- Uddannelse  1
- Arbejde  2
- Fritidsaktiviteter  3
- Børn  4
- Egen fornøjelse/underholdning  5
- Andet  6

9. Hvad låner du på biblioteket? Faglitteratur  1
- (Sæt kryds ved de vigtigste, max. tre krydser) Skønlitteratur  2
- Børnebøger  3
- Tidsskrifter  4
- Musik  5
- Videofilm/dvd  6
- Computerspil  7
- Andet  8

**Din brug af bibliotekaren på biblioteket:**

10. Bruger du bibliotekaren til:

- at fremfinde bøger eller andet materiale, du kender i forvejen Ja  1 Nej  2
- at få svar på om bøgerne eller materialet er hjemme Ja  1 Nej  2
- at undersøge om biblioteket har lignende bøger/materiale Ja  1 Nej  2
- at reservere bøger eller andet materiale Ja  1 Nej  2
- at fremfinde bøger eller andet materiale om et givent emne Ja  1 Nej  2

11. På en skala fra 0 til 5, hvor 0 er "aldrig" og 5 er "hver gang", hvor ofte benytter du dig af bibliotekaren?

Sæt ring om det tal du synes passer bedst:

<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
aldrig					Hver gang

**Hvor enig er du i følgende udsagn: (sæt kun ét kryds)**

12. Det er vigtigt for mig, at der er en bibliotekar til rådighed på biblioteket, som kan hjælpe med biblioteksfaglige spørgsmål

Helt enig 1	Delvis enig 2	Hverken Eller 3	Delvis uenig 4	Helt uenig 5	Ved ikke 9

**Din brug af bibliotekets hjemmeside:**

13. Hvor ofte bruger du bibliotekets hjemmeside.....  1 eller flere gange om ugen

2 1-3 gange om måneden

3 1-5 gange på et halvt år

4 1-2 gange om året

5 Aldrig

14. Hvor lang tid bruger du oftest pr. besøg.....  1 0 - ½ time

2 ½- 1 time

3 1 - 1½ time

4 1½- 2 timer

5 mere end 2 timer

15. Bruger du bibliotekets hjemmeside til:

at søge bøger eller andet materiale, du kender i forvejen

Ja  1 Nej  2

at få svar på om bøgerne eller materialet er hjemme

Ja  1 Nej  2

at undersøge om biblioteket har lignende bøger/materiale

Ja  1 Nej  2

at reservere bøger eller andet materiale

Ja  1 Nej  2

at søge bøger eller andet materiale om et givent emne

Ja  1 Nej  2

**Hvor enig er du i følgende udsagn: (sæt kun ét kryds)**

16. Det er vigtigt for mig, at biblioteket har en hjemmeside, hvor jeg selv har mulighed for at søge i bibliotekets database

Helt enig 1	Delvis enig 2	Hverken Eller 3	Delvis uenig 4	Helt uenig 5	Ved ikke 6

**Og til sidst:**

17. Kan du forestille dig et bibliotek uden bibliotekar?

Ja  1 Nej  2

**Tak for din hjælp**

### **Bilag 3**

Forskningsmetode er det første fag, hvor vi har stiftet bekendtskab med eksamensformen portfolio. Denne eksamensform lægger op til en helt anden type undervisning end vi ellers er vant til her på DB. Dette, kombineret med, at forskningsmetode og statistik for de fleste er helt nyt og svært stof virkede fra start meget overvældende. Man kunne med fordel have brugt noget mere tid, allerførst i forløbet, på introduktion af selve undervisningsforløbet og formen.

Det var svært at lave den første opgave så tidligt i forløbet, men vi kan tydeligt se vores egen udvikling gennem de fire opgaver, og vi har forsøgt at lære af vejledningerne, fra den ene opgave til den næste.

Helt konkret har vi lært meget af at konstruere et spørgeskema. Opstillingen og formen, samt måden hvorpå man indsamler svar. Vi havde stor succes med vores, men har da nye ideer til hvordan det kan gøres endnu bedre. Med P3 og P4 har vi fået et grundlæggende kendskab til, hvorledes man kan behandle data, selvom vi nok skal prøve det nogen gange endnu, før vi føler os helt sikre.

(fælles)