

Validitetsbetragtninger ved internetbaserede statistiske undersøgelser

Diskussion og analyse af validitet ved
internetbaserede statistiske undersøgelser.
Analyse er, hvor muligt, understøttet af faktiske
tal fra dette års internetundersøgelse foretaget af znail.

Maj 1999, Anders Blauenfeldt, **znail**

Kolofon

Sat i Verdana, Book Antiqua og Arial.
Reference til denne rapport
er tilladt med nøjagtig kildeangivelse.
Anders Blauenfeldt, znail.

Indholdsfortegnelse

Resumé	4
Forord	4
Indledning	5
Afgrensning.....	6
Metodeovervejelser	7
Stikprøvetagning	7
Internetbaserede undersøgelser.....	7
Analyse	10
Reklamering	10
Motivering af respondenter.....	14
Interaktion med respondenter	17
Konklusioner, kommunikation med respondenter	17
Test	19
Test for sandhed i besvarelsen	20
Internet som dataindsamlingsmedie	21
Tekniske muligheder og begrænsninger	21
Konklusioner, erfaringer.....	25
Afrunding, forslag til videre arbejde	27
Kombination af RDD og internet.....	27
International dataindsamling.....	27
Begrebsforklaringer	28
Bilag: selve spørgeskemaet	29

Resumé

Denne rapport søger at analysere og diskutere validitetsproblemer vedrørende internetbaseret statistisk dataindsamling. Ydermere indeholder rapporten en generel diskussion omkring denne undersøgelsesform, samt en konkret gennemgang af de forskellige aspekter ved at køre en internetundersøgelse.

Keywords:

Internet, validitet, demografi , nettet, statistik, web.

Forord

znail har i 1998 og 1999 foretaget internetbaserede undersøgelser af danskeres internetvaner. Disse undersøgelser er publiceret via znail.dk/survey - og kan downloades gratis.

Nærværende rapport er udarbejdet af **znail** til Forskningsministeriet. Rapporten kan betegnes som et validitetstillæg til analyserapporten '*Danskeres internetvaner 1999*'. For det fulde udbytte af diskussionen og konklusionerne i nærværende rapport bør ovenstående undersøgelsesrapport læses. '*Danskeres internetvaner 1999*' kan downloades gratis fra znail.dk/survey.

Tak til cand. polit Søren Baunsgaard for gode kommentarer og kritisk korrektur og til Bjarne Ersbøll, IMM ved Danmarks Tekniske Universitet for gode råd og vejledning.

Indledning

Over 1 million mennesker i Danmark bruger internet. Over 150 millioner bruger internet på verdensplan¹. Mediet er i hurtig fremdrift og man oplever en bred opfattelse af at internettet vil være udbredt i den brede befolkning indenfor få år.

Denne hurtige udbredelse af et helt nyt medie/kommunikationsværktøj kan måske bedst beskrives som et paradigmeskift, indenfor områderne medie og kommunikation. Således opleves i øjeblikket hvorledes brancher og arbejdsområder i disse år berøres af udviklingen indenfor internet, nogle dramatisk, nogle mindre dramatisk, men stort set alle.

Blandt meget andet har internettet introduceret en ny dataindsamlingsmetode som konkurrent til eksisterende dataindsamlingsmetoder: telefoninterview, personligt interview eller forskellige former for papirbaserede spørgeskemaer. Dette område, internetbaseret statistisk dataindsamling, er fokus for denne rapport.

Vi vil her se på hvor god en dataindsamlingsmetode internettet kan give os. Hvilke faktorer spiller ind i forbindelse med internetbaserede undersøgelser. Hvorved opstår bias - hvilke usikkerhedsmomenter kan umiddelbart elimineres og hvilke kan måske ikke uden videre elimineres.

Vi vil beskrive internetundersøgelser generelt samt, med udgangspunkt i den netop afsluttede "*anden årlige store danske internetundersøgelse*" - fremover benævnt [znail99], fremdrage konkrete eksempler på internetbaseret dataindsamling og mulige problemer herved. Disse konkrete eksempler, case-stories, vil så vidt muligt være afgrænset i grå bokse, og kan således nemt overskues.

Da dette undersøgelsesmedie er nyt findes der meget få metodemæssige studier vi kan underbygge denne rapport med. Således skal denne rapport ikke ses som en konkluderende undersøgelse og analyse, men nærmere som et diskussionsoplæg og en problematisering af hvillke problemer man bør være sig bevidste i forbindelse med internetbaserede undersøgelser.

Vi forudsætter i denne rapport et vis kendskab til internet og relaterede emner. Således vil vi bruge begreber som søgemaskiner, bannerreklamer osv og forudsætte at læseren på forhånd har kendskab hertil. Bagest i denne rapport findes dog en opsummering af nogle almindelige begreber. Internetteknologien

¹ Det irske firma NUA har de seneste år samlet informationer omkring antallet af personer online på verdensplan. Som de selv gør opmærksom på er denne estimation utrolig svær og som oftest behæftet med stor usikkerhed. Alligevel er det dog interessant at følge udviklingen på www.nua.com.

har introduceret mange nye begreber i vores hverdag, og en meget vigtig forudsætning for at diskutere dette område er derfor en fælles opfattelse af begreberne.

Afgrænsning

Vi vil i denne rapport afgrænse os til at se på generelle undersøgelser foretaget med internettet som indsamlingsmedie. Ved specifikke undersøgelser - fx et firma der ønsker at spørge sine kunder omkring produktbrug eller lignende, kan diskussionen omkring brug af internet som indsamlingsmedie være en ganske anden. Såfremt kundesegmentet må antages at kunne bruge internet, kan det være en oplagt og billig dataindsamlingsmetode, men hvis ikke kundesegmentet bruger internet er det naturligvis en dårlig ide.

Metodeovervejelser

Stikprøvetagning

Statistiske undersøgelser der søger at beskrive udvalgte forhold eller vaner hos en stor gruppe mennesker - eksempelvis indkøbsvaner eller politiske tilhørsforhold - baserer sig som oftest på udsagn fra et uddrag af gruppen. Hvorledes dette uddrag af befolkningen, stikprøven, foretages bestemmer ofte kvaliteten af undersøgelsen.

Der skelnes mellem *tilfældig* stikprøvetagning og *ikke-tilfældig* stikprøvetagning, hvor ikke-tilfældig stikprøvetagning kan have forskellige grader af tilfældighed. Hvis stikprøveudtagningen favoriserer nogle grupper af hele mængden fremfor andre opstår en skævhed i undersøgelsen - undersøgelsen er *biased*.

Tænk på en (stor) skål med 100.000 bolde i 5 forskellige farver ligeligt fordelt mellem farverne. Vi kender ikke fordelingen og vil gerne finde den ved hjælp af en stikprøve. Vi blander boldene i skålen, udtrækker derefter et antal bolde og noterer farverne i vores stikprøve. Hvis nu de grønne bolde, uden vort kendskab, var meget tungere end alle de andre, og derfor havde lagt sig nederst i skålen ville vores måling ikke være specielt god, om så vi udtrak 80.000 ud af de 100.000 bolde. Målingen var *biased* væk fra de grønne bolde. Hvis derimod sandsynligheden for at udtrække de 5 forskellige farver var lige stor, behøvede vi kun en mindre måling før vi kunne bestemme fordelingen af farver. Man kan vise statistisk, at et udtræk på omkring 1200 i de fleste tilfælde er passende såfremt de nøjagtigt repræsenterer den samlede mængde. Dette er således også ca. det antal mennesker analyseinstitutter adspørger ved fx politiske meningsmålinger.

Selvudvælgelse

En ofte benyttet metode til undersøgelser gældende for en hel befolkning er RDD (random digit dialing) - en computer sammensætter et tilfældigt telefonnummer og ringer op. Denne metode kan dog indeholde det problem, at deltagerne kan vælge at sige nej til at deltage. Derved opstår et problem vedrørende selvudvælgelse og vi kan ikke være sikre på at gruppen af deltagere er 100% repræsentativ for hele befolkningen. Problemet kan beskrives som: *vi foretager ikke en analyse blandt alle, men derimod blandt gruppen af personer der vælger at sige ja til at deltage i undersøgelsen.*

Internetbaserede undersøgelser

Selvudvælgelsesproblematikken bliver tydeligere hvis vi vil foretage generelle undersøgelser via internet. Et stort problem er at: **vi kender ikke hele mængden af**

internetbrugere. Der findes ikke et centralt register af internetbrugere eller email-adresser, og det er derfor praktisk umuligt at foretage en tilfældig generering af personer svarende til RDD, beskrevet ovenfor. Vi er nødt til at bruge selvudvælgelse i endnu højere grad ved, i stedet for at kontakte vore respondenter, så forsøge at "lokke" dem til vores undersøgelse ved eksempelvis at lade dem deltage i en lodtrækning. Dette kan gøre vores usikkerhed større, da vi her er nødt til at henvende os til en noget større mængde, sammenlignet med fx telefoninterviews, for at få samme antal respondenter. Problemet kan her beskrives som: *vi foretager ikke en analyse blandt alle internetbrugere, men blandt gruppen af personer der ser/hører om undersøgelsen og derefter selv tager initiativ til at deltage i undersøgelsen.*

Forskellige forhold er af betydning for om en internetbruger deltager i en internetbaseret undersøgelse. Man skal eksempelvis have tid til at deltage, man skal kunne finde ud af at deltage, man skal ikke give op på halvvejen, - eller kort formuleret:

Man skal have lyst, og evne til at deltage

Det er logisk at argumentere for følgende: bliver man ringet op og adspurgt af et menneske om man vil deltage i en undersøgelse og svare på et par spørgsmål er det sværere at sige "*nej jeg har ikke tid*", end hvis man blev ringet op af en computerstemme der siger: *tryk 1 for at deltage, tryk 2 for at afslutte*. Der findes selvfølgelig personer der vil deltage alligvel. Vi kunne kalde denne gruppe: *den altruistiske respondentgruppe* - folk der deltager af ren og skær lyst, og for at hjælpe undersøgelsen på vej. Men for at alle respondenter ikke skal være personer der blot har tid og lyst til at deltage - hvilket sandsynligvis ej heller vil give respondenter nok til et statistisk grundlag - kan vi bruge en lodtrækning blandt de deltagende for at give noget igen. Man giver altså respondenter et lod i en lodtrækning mod at besvare nogle spørgsmål, og derved skabes en mere generel lyst til at deltage.

Endeligt, efter at respondenter er blevet ført til undersøgelsen, accepteret betingelserne og påbegyndt besvarelsen, er det vigtigt at vi også sørger for at en så stor del som muligt af respondenterne gennemfører besvarelsen.

Vi kan opstille følgende generelle betragtninger i forbindelse med besvarelse af et internetbaseret spørgeskema

- en smule interneterfaring kræves for overhovedet at udfylde det web-baserede spørgeskema

- folk med mere erfaring på nettet må anses for at være mere motiverede til at fortælle om deres glæder og frustrationer ved internet end folk der kun har brugt internet et lille stykke tid eller stadig er meget usikre på denne nye teknologi.
- spørgeskemaet vil som oftest henvende sig sprogligt til voksne mennesker

Disse generelle betragtninger tyder alle på at vi vil opleve en overvægt af erfarne internetbrugere blandt respondenterne i en internetundersøgelse.

Lad os, før vi i næste afsnit uddyber disse emner, opsummere vore hidtidige betragtninger angående kontakt til respondenter. Vi kan, kronologisk i forbindelse med deltagelsen, herunder opstille de forskellige faser ved kommunikation med respondenter som:



Analyse

Vi vil i dette afsnit gennemgå de, i det foregående afsnit, fremtrukne emner i forbindelse med kontakt til respondenter ved en internetbaseret undersøgelse.

Reklamering

På internet har vi de sidste år, i takt med en udpræget kommercialisering af dette medie, oplevet udbredelsen af banner-reklamer. Bannerreklamer er således nu de facto standard for reklamering på nettet. Der findes forskellige bannertyper, men fælles for dem alle er at de består af et lille rektangulært grafikbillede, som oftest placeret øverst på en webside. En oplagt måde at henvende sig til sine potentielle respondenter blandt internetbrugere, er derfor at bruge denne reklameform.

Forestiller vi os, i den forbindelse, at vi på Rådhuspladsen i København spørger alle der kommer forbi hvor de bor, er det ikke svært at indse at svaret ikke vil være repræsentativ for hele Danmarks befolkning. Sådan er det også når vi kommunikerer via internet. Forestiller vi os således at vi placerer en reklame til vores undersøgelse på et websted der handler med edb-udstyr, henter vi således vores respondentgruppe fra personer der bruger internet til at handle edb-udstyr - med overvejende sandsynlighed vil dette give en bias, fx i kønsfordelingen i retning af flere mænd. Det er med andre ord væsentligt hvor man reklamerer for sin undersøgelse.

Under den forudsætning at man, i sin respondentgruppe, søger den maksimale repræsentativitet for hele mængden af internetbrugere, vil man naturligt være interesseret i at reklamere så bredt og objektivt som muligt. Her vil det være naturligt at vælge de mest brede og benyttede internettjenester - fx søgemaskiner. Da de fleste søgesteder tilbyder direkte markedsføring i forbindelse med bannerplaceringer - ved søgning på ordet blomst, bliver en reklame for InterFlora fremvist - er det her igen vigtigt at placeringen af banneret bliver neutralt fordelt for søgestedet.

I forbindelse med internetmarkedsføring skal man, afhængigt af bannerets signalværdi, fremvise et reklamebanner mellem 100 og helt op til 500 gange før der gennemsnitligt er én bruger der klikker på banneret². Antager vi en click-thru rate³

² Man mener at den gennemsnitlige rate af klik per banner på det danske del af internet ligger mellem 1% og 1,5%. Dette varierer dog meget - bank og forsikringskampagner kan ligge under 0,5% mens konkurrencer og lignende kan ligge langt højere.

³ click-thru måles som regel i promille. 10 promille, altså 10 klik for hver 1000 fremvisninger, anses for en høj click-thru rate. Dette afhænger dog meget af stedet man reklamerer og teknologien der ligger bag fremvisningen af reklamer. Enkelte internetsteder er allerede nu så avancerede at de kan styre fuldstændigt hvor mange gange et banner skal vises

på 10 klik per 1000 fremvisninger, og vil vi gerne have 5000 besøgende til vores undersøgelse, skal vi således ud at investere i 500.000 bannereksponeeringer. Priser på bannerreklamer varierer meget og vi skal derfor ikke her opstille budgetter for denne reklameringsudgift. Det er yderligere, i en analysesituation, ikke usandsynligt at man kan lave aftaler med forskellige internetsteder omkring udveksling af data mod gratis bannerreklame.

Tabel 1 - bannereksponeeringer

eksponeringer	293.176
clicks	2.617
clicks/1000	8,9

Tabel 2 - hvorfra hørt om undersøgelsen ?

hørt om undersøgelsen:	n	%
andet	1297	43,06%
kønsfordeling	k/m: 41%/59%	
hørt fra venner og bekendte	268	8,90%
kønsfordeling	k/m: 40%/60%	
klikket på en bannerreklame	1228	40,77%
kønsfordeling	k/m: 52%/48%	
læst i aviser/blade	21	0,70%
kønsfordeling	k/m: 33%/67%	
læst i en newsgroup	148	4,91%
kønsfordeling	k/m: 35%/65%	
ej besvaret	50	1,66%

Tabel 3 - genganger ?

deltog sidste år	448	14,87%
deltog ikke sidste år	2528	83,93%
ej besvaret	36	1,20%

Boks 1 - Hvorledes blev respondenterne ført til undersøgelsen, [znail99].

Som det kan læses i rapporten *Danskeres internetvaner 1999* reklameredes der, ved dette års undersøgelse, ved hjælp af banner-reklamer på 5 af de meget besøgte, og generelle, danske internetsteder; yahoo.dk, eon.dk, krak.dk, tv2.dk og jubii.dk.

Tabel 1 viser os de samlede tal fra disse bannerkampagner og vi ser at 293.176 fremvisninger af vores reklame-banner har medført 2.617 besøgende på undersøgelses-sitet. Da vi nåede op på omkring 4.500 påbegyndte besvarelser og heraf 3.012 valide gennemførte besvarelser, må der med andre ord være en del respondenter der er kommet til undersøgelsen *uden* at have klikket på vores reklamebanner. For at undersøge dette spurgte vi til slut i undersøgelsen vore respondenter hvor man havde hørt om undersøgelsen. I Tabel 2 ses at 41% af vore respondenter er kommet til undersøgelsen via banner-reklamen, men at en ligeså stor andel (43%) har markeret 'andet'. I forbindelse med at undersøgelsen sattes igang rundsendte vi en mail til en mailingliste baseret på sidste års deltagere, og derfor kan vi formode at de respondenter der deltog grundet denne mail har valgt 'andet'. Dette kan vi dog igen underbygge med data fra undersøgelsen da vi ydermere spurgte om respondenteren også deltog sidste år. Vi ser i Tabel 3 at 448 respondenter er gengangere. Af disse 448 respondenter må vi dog antage at der stadig findes en mængde der er ført til undersøgelsen via et banner - og altså ikke deltager pga en mail fra os. Således har 76 respondenter deltaget i undersøgelsen sidste år, men stadig klikket på et banner i år. For at undersøge en eventuel bias i hvordan respondenteren er ført til undersøgelsen ser vi i Tabel 2 ligeledes kønsfordelingerne. Vi ser en tendens til at relativt flest kvinder er kommet til undersøgelsen ved at klikke på et reklamebanner, hvorimod fx avertering i newsgroups vil tiltrække flest mænd. Med andre ord kan reklameringsform give bias.

Opsummering

- 2617 af de ialt 4500 påbegyndte besvarelser kom til undersøgelsen via en banner-reklame.
- Under halvdelen af de 3012 valide besvarelser har markeret at de kom til undersøgelsen via et banner.
- Næsten 10 % har markeret 'hørt fra venner og bekendte'
- En stor gruppe på 43% har markeret 'andet'.
- Vi observerer en forskel i kønsfordelingen afhængigt af hvorledes man har hørt om undersøgelsen.

Vi ser i Boks 1 at respondenter der føres til en internetundersøgelse via et banner tilsyneladende kan have en overvægt af kvinder, og at respondenter der bliver gjort opmærksom på undersøgelsen via en news-group har en overvægt af mænd.

per bruger - målt ved IP-nummeret på maskine. Således kan et banner fremvises indtil en bruger klikker, eller man kan fremvise det fx 5 gange og så "give op".

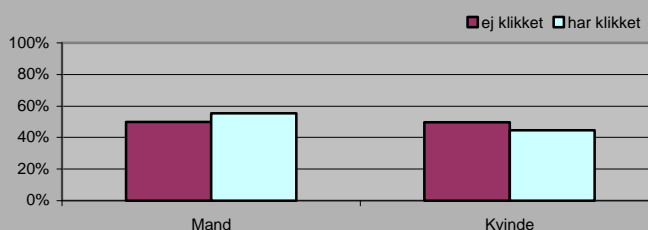
Et meget væsentligt spørgsmål i forbindelse med bannerreklamering for en internetundersøgelse er, om det kan være et problem at folk kun kommer til undersøgelsen via bannerreklamer. Ved denne form for tiltrækning af respondenter, foretages jo udelukkende en undersøgelse blandt den del af internetbrugere der rent faktisk klikker på bannerreklamer. Og vi kan ikke være sikre på at denne delmængde er repræsentativ for hele mængden af internetbrugere. Som det ses i Boks 2 kan vi antage at omkring en femtedel af internetbrugere ikke har prøvet at klikke på et reklamebanner. Denne gruppe har, som det ses, en relativ overvægt af nybegyndere, kvinder og personer under 20 år henholdsvis over 40 år. Kommer respondenterne til undersøgelsen udelukkende via et banner kan man derfor risikere en bias væk fra denne mængde.

Boks 2 - Hvem har klikket på et banner?, [znail99]

2375 respondenter markerede at de har prøvet at klikke på et banner. Lad os her sammenligne denne gruppe med de 637 der ikke har klikket på et banner.

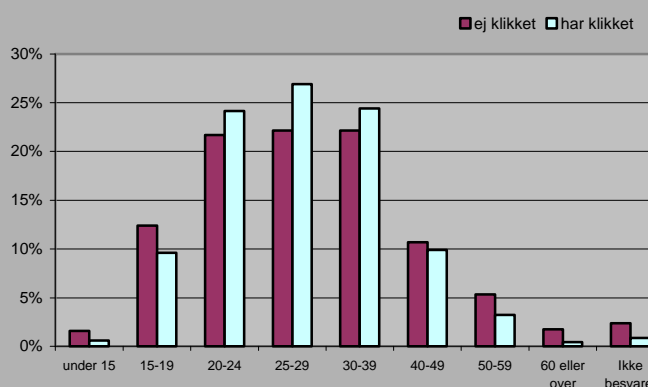
Tabel 4 - kønsfordeling

	Ej klikket		Klikket	
	n	%	n	%
Mand	318	50,0%	1316	55,4%
Kvinde	317	49,8%	1059	44,6%
Ikke besvaret	2	0,3%	0	0,0%



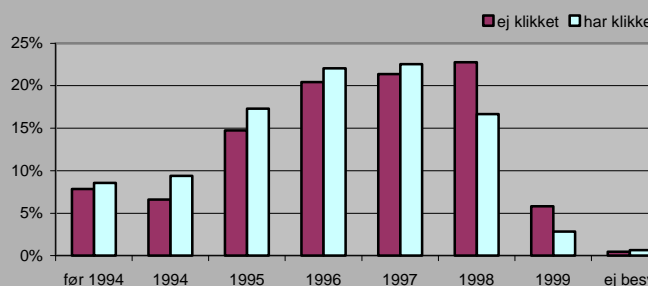
Tabel 5 - aldersfordeling

	Ej klikket		Klikket	
	n	%	n	%
under 15	10	1,6%	14	0,6%
15-19	79	12,4%	228	9,6%
20-24	138	21,7%	573	24,1%
25-29	141	22,1%	639	26,9%
30-39	141	22,1%	579	24,4%
40-49	68	10,7%	235	9,9%
50-59	34	5,3%	77	3,2%
60 eller over	11	1,7%	10	0,4%
Ikke besvaret	15	2,4%	20	0,8%



Tabel 6 - web-erfaring

	Ej klikket		Klikket	
	n	%	n	%
før 1994	50	7,85%	203	8,6%
1994	42	6,59%	223	9,4%
1995	94	14,76%	411	17,3%
1996	130	20,41%	523	22,0%
1997	136	21,35%	535	22,5%
1998	145	22,76%	396	16,7%
1999	37	5,81%	68	2,9%
ej besv	3	0,47%	16	0,7%



Alle ovenstående tabeller er testet med Fishers two-tailed test for signifikans. Ved alle tre områder kan vi konkludere at der er signifikant forskellighed i målingerne.

Man bør dog være opmærksom på en mulig forskel i at "jeg har på et tidspunkt klikket på et banner" og "jeg lader mig påvirke af et banner". Denne problematik kan muligvis forklare det modstridende i at vi fra Boks 2 må forvente en overvægt af mænd blandt respondenterne der kommer til undersøgelsen via et banner, men at vi i Boks 1 observerer en overvægt af kvinder.

I forbindelse med ovenstående, og med modsat fortegn, kan vi ligeledes stille spørgsmålet: *Kan der være et problem ved at alle kan deltage i undersøgelsen uanset måden hvorpå de er blevet gjort opmærksom på undersøgelsen?*

Svaret herpå er naturligvis "ja", og kan forklares med følgende eksempel: en 75-årig person ser en bannerreklame for vores undersøgelse og deltager i undersøgelsen. Vedkommende vil gerne hjælpe undersøgelsen på vej, og distribuerer derfor et link til undersøgelsen overfor venner og bekendte - også i aldersgruppen over 70 år. Dette vil give en bias til fordel for gruppen af ældre mennesker idet én bannerfremvisning overfor én 75-årig medfører en hel mængde af besvarelser fra samme aldersgruppe. Et tænkt eksempel, men den grundlæggende problemstilling bag er væsentlig og kan naturligvis overføres på bias i køns-fordeling, bopælsfordeling osv. Som en uddybning af denne problemstilling kan vi se på et andet eksempel: man kunne forestille sig at undersøgelsen blev omtalt i ugebladet "Alt for damerne" med en notits og en opfordring til at deltage - og således ville en bias imod flere kvinder være skabt.

Ovenstående kan sammentrækkes i to problemstillinger; (1:) *gruppen af internetbrugere der klikker på bannerreklamer er ikke repræsentative for hele mængden af internetbrugere* og (2:) *ved at lade alle deltage i undersøgelsen kan vi risikere uforudsete ikke umiddelbart målelige fluktuationer.*

En vigtig forskel på disse problemer er at problemstillingen (1) ikke vil variere med antallet af respondenter, hvorimod problemstillingen (2) vil blive mindre jo flere respondenter vi bygger vores undersøgelse på. En undersøgelse med fx 50.000 respondenter er således følsom overfor problemstillingen (1), men temmelig robust overfor problemstillingen (2), hvorimod en undersøgelse med kun 500 respondenter vil være følsom overfor begge problemstillinger (1) og (2).

I forbindelse med begrænsning af disse problemstillinger er det vigtigt at være sig bevidst hvilke tekniske muligheder der ligger i forbindelse med internetbaserede undersøgelser. Fra et ethvert http-request (forespørgsel om en internetside) kan man udtrække en 'referer' - beskrivelse om hvor internetbrugeren befandt sig da vedkommende foretog forespørgslen. Man kan således se hvorfra internetbrugeren kommer. Dette bruges ofte aktivt i forbindelse med logfilsanalyser, og kan i dette tilfælde, med den rette tekniske opsætning, gøre os i stand til at udskille de forskellige mængder af respondenter. Mængden der kommer via hvilke bannere

og naturligvis mængden der ikke kommer til undersøgelsen via et banner. Man bør derfor gøre sig tekniske overvejelser til at kunne løse dette problem i forbindelse med udvikling af internetbaserede spørgeskemaer.

Delkonklusion

I forbindelse med annoncering for en undersøgelse på nettet kan vi konkludere at der ved bannerreklamering bør korrigeres for en relativ overvægt af mænd, 20-40 årige og erfarne internetbrugere, blandt de internetbrugere der har prøvet at klikke på et banner. Dog observeres (Boks 1) det modsatte, idet vi ved [znail99] observerer en relativ overvægt af kvinder i gruppen af respondenter der er kommet til undersøgelsen via et banner.

Ligeledes må vi, specielt ved undersøgelser med forholdsvis få respondenter, forvente en vis risiko for umålbare fluktuationer ved at lade alle deltage i undersøgelsen.

Motivering af respondenter

Klik og vind. Så simpelt kan det som regel formuleres hvis man skal give opskriften på hvorledes man hurtigt kan lokke en stor gruppe internetbrugere mod et internetsted. Som beskrevet tidligere i det generelle afsnit, kan vi bruge en lodtrækning til at give noget igen, for det besvær respondenterne har ved at udfylde skemaet.

Man kunne dog forestille sig alternative "aflønningsmetoder" end et lod til en lodtrækning. Fx kunne man slet og ret give respondenterne løn for at deltage. Dette vil dog hurtigt andrage helt andre beløbsstørrelser. Hvis vi ser i Boks 3 androg den samlede præmiepulje ved dette års znail-undersøgelse under 3,5 kroner per respondenter - og for 3,5 kroner må man formode at det er svært at lokke folk til at deltage. En andet problemstilling herved ligger naturligvis i overdragelsen af beløbet til respondenterne - idet vi ikke har fysisk kontakt og derved ikke kan overdrage kontanter. Sålænge internetbaseret valuta ikke er udbredt vil dette altså udgøre et problem. Man kunne dog forestille sig en digital gavekortsbetalingsform - fx et gavekort på et par dollars til internetboghandleren amazon.com⁴ eller lignende, men igen vil beløbsstørrelsen nok være en barriere, da det er vores fornemmelse at man hellere vil have et lod og måske vinde en stor præmie end at modtage fx 10 kroner.

I forbindelse med en lodtrækning er det naturligvis vigtigt at lodtrækningspræmien ikke genererer en bias - fx ville deltagelse i lodtrækningen om et lystfisker-sæt nok give en bias mod relativt flere mandlige respondenter end

⁴ Forskellige internetsteder har på det seneste foræret gavekort væk til amazon.com. Således kan man få et par dollars rabat ved næste bogkøb ved at taste en kode ind man har fået fremsendt.

kvindelige, ligesom en lodtrækning om en 'sony playstation' vil henvende sig relativt mest til unge/børn. Den mest objektive præmie man umiddelbart kan forestille sig, som ikke vil skabe en bias, vil være rede penge.

Tabel 7 - hvor tit deltager du i konkurrencer via internettet ?

alle dem jeg kan finde	308	10,2%
ofte - dem jeg støder på en gang imellem	1020	33,9%
sjældent	659	21,9%
aldrig	142	4,7%
ikke besvaret	51	1,7%
ialt	3012	100,0%

Boks 3 - lodtrækningspræmier, [znail99]

Vi har i vore undersøgelser valgt at udlove rejsepræmier, ud fra den betragtning at et rejsegavekort kan stort set alle bruge - enten til sig selv eller venner/familie. Vi er dog bevidste om argumentet for, at rede penge skulle være den bedste og mest objektive præmie - men heri ligger et væsentligt problem i at det vil være svært at få en sponsor til at sponsorere en sådan præmie. Herudover er det vores postulat at en rejsepræmie genererer flere respondenter en en "kedelig" pengepræmie. Således kunne vi i år trække lod om en hovedpræmie, rejsegavekort på 5000 kroner og 10 andenpræmier - rejsegavekort på 500 kroner.

Da det på "nettet" efterhånden er ganske udbredt at have lodtrækninger på forskellige web-steder, er der derved også skabt en gruppe internetbrugere der deltager i alle de konkurrencer/lodtrækninger de kan komme i nærheden af. Hvis vores respondentgruppe udelukkende bestod af denne type internetbruger, ville vi igen stå overfor en mulig bias. For at undersøge dette punkt lidt nærmere placerede vi til sidst i undersøgelsen et par spørgsmål omkring selve deltagelsen, blandt andet med spørgsmålet: *hvor tit deltager du i konkurrencer på nettet?*

Vi mener, at Tabel 7 viser os at vores respondentgruppe ikke udgøres af "konkurrence-ryttere" - altså folk der deltager i mange konkurrencer på "nettet". 1/4 af respondenter deltager sjældent eller aldrig i konkurrencer på nettet og kun 10% har svaret "alle dem jeg kan finde". Hvorvidt lige netop svaret på dette spørgsmål så er repræsentativt for hele gruppen af danske internetbrugere, er svært at svare på, men vi er tilfredse med fordelingen og mener ikke den udgør en væsentlig bias-risiko. At folk udelukkende skulle deltage i undersøgelsen for at deltage i lodtrækningen kan vi yderligere modargumentere ved at alle har givet sig god tid til at besvare spørgeskemaet. Spørgeskemaet bestod af 8 sider og vi målte tiden fra man startede på den første side til man afleverede spørgeskemaet. Således havde vi også mulighed for at luge ud i de eventuelle tomme besvarelser - personer der kun havde udfyldt navn og email og intet andet (det drejede sig dog kun om ialt 3 besvarelser!). Den gennemsnitlige tid brugt på skemaet er 14 minutter og 25 sekunder 99,4% af respondenterne brugte mere end 6 minutter på besvarelsen og 77,5% af respondenterne brugte mere end 10 minutter. Med andre ord har stort set alle givet sig god tid til besvarelsen, hvilket også giver en indikation af at besvarelserne er "rigtige" - at folk faktisk har givet os det sande billede af deres internetbrug.

Note: det sande gennemsnit af tidsforbrugeren ved alle besvarelser er noget større, men her fraregner vi de godt 100 respondenter hvor besvarelsen har taget over 1 time - hvis man bruger 7 timer på spørgeskemaet formoder vi, det er fordi man laver noget andet i mellemtiden og derfor er det ikke 7 timers effektiv spørgeskema-udfyldning

Vi kan yderligere diskutere udlovning af lodtrækningspræmier med følgende postulater:

- unge mennesker (studerende) har mere brug for en præmie på eksempelvis 10000 kr og dermed også mere lyst til at deltage i undersøgelsen.

- erfarne internetbrugere er "vant" til at se reklamebannere med løfter om gratis glæder og store lodtrækningspræmier, og lader sig dermed ikke så nemt overtale til at deltage.

Andet postulat modsiger dog vores konklusion fra sidste afsnit om at reklamering ved hjælp af et banner kan skabe en bias mod erfarne internetbrugere.

Delkonklusion

I forbindelse med motivering af respondenter vil den mest oplagte fremgangsmåde være at foretage en lodtrækning mellem respondenterne. Lodtrækningspræmien skal vælges så neutral som muligt - eksempelvis pengepræmier.

Boks 4 - Gruppen af respondenter der ikke gennemførte, [znail99]

1525 påbegyndte besvarelser blev aldrig afleveret. Størstedelen opgav allerede på første side men vi har dog omkring 450 personer der nåede at taste de grundlæggende oplysninger ind før de opgav midt i undersøgelsen. Vi må antage at understående tal er behæftet med stor usikkerhed idet årsager til at stoppe midt i undersøgelsen kan være mange og at vi samtidig ikke her tester for om personen både har afleveret en halvt færdig og så måske senere hen en helt færdig besvarelse.

Tabel 8 - kønsfordeling, ej færdiggjort spg-skema

	Ej færdiggjort		Færdiggjort	
	n	%	n	%
Mand	204	47,78%	1634	54,25%
Kvinde	223	52,22%	1376	45,68%

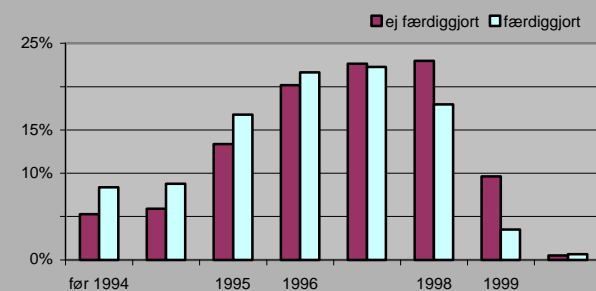
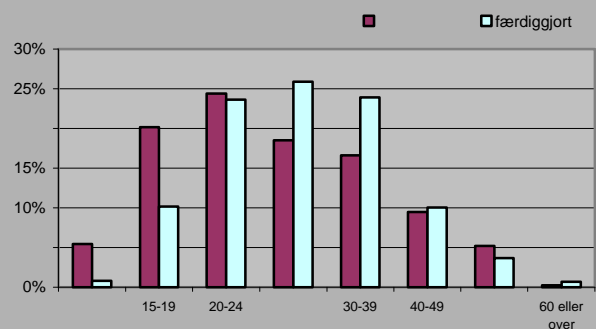
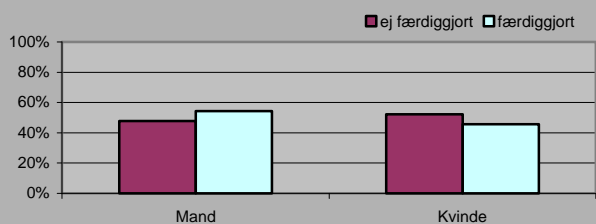
Tabel 9 - aldersfordeling, ej færdiggjort spg-skema

	Ej færdiggjort		Færdiggjort	
	n	%	n	%
under 15	23	5,45%	24	0,80%
15-19	85	20,14%	307	10,19%
20-24	103	24,41%	711	23,61%
25-29	78	18,48%	780	25,90%
30-39	70	16,59%	720	23,90%
40-49	40	9,48%	303	10,06%
50-59	22	5,21%	111	3,69%
60 eller over	1	0,24%	21	0,70%

Tabel 10 - weberfaring, ej færdiggjort spg-skema

	Ej færdiggjort		Færdiggjort	
	n	%	n	%
før 1994	17	5,28%	253	8,40%
1994	19	5,90%	265	8,80%
1995	43	13,35%	505	16,77%
1996	65	20,19%	653	21,68%
1997	73	22,67%	671	22,28%
1998	74	22,98%	541	17,96%
1999	31	9,63%	105	3,49%

Alle ovenstående tabeller er testet med Fishers two-tailed test for signifikans. Ved alle tre områder kan vi konkludere at der er signifikant forskellighed i målingerne.



Interaktion med respondenter

I forbindelse med udfyldning af selve spørgeskemaet kan vi forestille os endnu en risiko for bias. Hvis eksempelvis spørgeskemaet er formuleret i et meget teknisk sprog eller et generelt meget svært sprog, vil vi med sandsynlighed miste mange børn og unge selvom vi får dem til at starte på spørgeskemaet. Hvis spørgeskemaet er meget langt og tager over en time at udfylde vil vi sandsynligvis igen opleve at miste mange respondenter undervejs.

Vi ser i Boks 4 en gennemgang af de respondenter der gav op på halvvejen ved den store danske internetundersøgelse 1999. Tallene er naturligvis meget usikre, men vi kan dog observere tendenser. Vi ser, som ovenfor forudssagt, en tendens til at der er større sandsynlighed for at unge personer giver op end ældre personer. Specielt gruppen af personer under 20 år. Ligeledes ses en tendens til at nybegyndere på nettet giver op på halvvejen. Disse observationer understreger vores generelle opfattelse af at vor gruppe af respondenter har en bias mod de mere erfarne internetbrugere, og dette bør man korigere for eller som minimum gøre opmærksom på i forbindelse med internetbaserede undersøgelser.

Konklusioner, kommunikation med respondenter

Vi må generelt konkludere at der ved de forskellige kommunikationsmomenter med respondenter; reklamering, motivering og interaktion, alle steder kan forekomme en uønsket frasortering af respondenter - og derved bias.

I forbindelse med annoncering for en undersøgelse er vi i en situation med modstridende målinger. Vi kan måle en overvægt af mænd, 20-40 årige samt internet-erfarne i gruppen af internetbrugere der har prøvet at klikke på et banner. Samtidig kan vi måle en overvægt af kvinder i gruppen af respondenter der er kommet til [znail99] via et banner (bygget på eget udsagn).

Der argumenteres for at man, specielt ved undersøgelser med forholdsvis få respondenter, bør være bevidst om en vis risiko for umålbare fluktuationer ved at lade alle deltage i undersøgelsen.

I forbindelse med motivering af respondenter kan vi konkludere at lodtrækningspræmier er den eneste reelle mulighed i øjeblikket. Valg af lodtrækningspræmie er vigtig og kan skabe en bias ved forkert valg. Vi anbefaler pengepræmier.

Ved interaktion med respondenter, selve udfyldelsen af spørgeskemaet, kan vi respondenter og uerfarne internetbrugere, giver op midt i besvarelsen.

Sammentrækkes de ovenstående punkter, må vi konkludere en generel bias til

Test

Som beskrevet tidligere ligger et fundamentalt problem ved internetundersøgelser i at man ikke kender hele mængden af internetbrugere, og derfor ikke kan sammenligne vore respondents repræsentativitet ved fx at sammenholde kønsfordeling, aldersfordeling, bopælsfordeling etc, med den 'ægte' fordeling. Det er således ganske få tests vi kan foretage i forbindelse med repræsentativiteten af vore respondenter, men vi vil i dette afsnit prøve at se på hvilke mulige tests der foreligger.

Test for repræsentativitet

En objektiv måling kan være browserfordelingen blandt vore respondenter. Med den rette tekniske opsætning kan vi måle hvilken browser hver enkel respondent bruger i forbindelse med udfyldningen af spørgeskemaet. Igen er vi ikke i stand til at sammenligne med den 'sande' fordeling da denne ikke vides og naturligt også ændres konstant. Et god tilnærmelse ville dog være at sammenligne vores måling med fx et gennemsnit af de 10 mest besøgte internetsteder i Danmark. Vi ser i Boks 5 at denne måling kan understrege tendensen til at respondenter ved internetundersøgelser tilhører den erfarne ende af brugerne.

Boks 5 - Browserfordeling, [znail99]

Tabel 11 - browserfordeling, jubii.dk sammenlignet med vore respondenter

	Explorer 5	Explorer 4	Explorer 3	Netscape 4	Netscape 3	Netscape 2
Jubii.dk		57 %	16%	19%	6%	1%
DSDI99	2,6%	59,8%	11,3%	22,3%	3,7%	0,3%

kilde: JAN (Jubii Annoncører Nyt),
DSDI99 = Den store danske internetundersøgelse 99

Vi ser for det første, at vore målinger stemmer relativt pænt overens. Det er positivt, og understreger igen at vor måling er forholdsvis repræsentativ for hele den danske internetbefolkning. Dog observerer vi en generel overvægt hos vore respondenter for brug af nyere browsers. Dette kan skyldes at tallene fra Jubii er fra januar måned (marts måneds tal kunne ej oplyses) og da der konstant opleves en strømning mod brug af nye browsere vil vore tal fra marts naturligt være en smule mere mod de nye versioner. Men det kan også skyldes at respondenterne har en bias mod de erfarne internetbrugere.

Ideer til andre objektive målinger kunne være at spørge respondenterne omkring hvilke konkrete internetsteder de har besøgt det sidste stykke tid og så sammenligne med faktuelle besøgstal fra disse internetsteder.

Eller at sammenligne respondenternes emailadresser med det faktiske kundefordeling hos de danske internetudbydere.
Fælles for disse målinger vil dog være en stor usikkerhed.

Test for sandhed i besvarelsen

Hvor meget kan vi stole på besvarelsene i en internetbaseret undersøgelse? Kan man argumentere for en forskel i sandheden i besvarelsen sammenlignet med traditionelle udspørgningsmetoder? Kan man forestille sig måder at teste denne sandhed?

Som tidligere, med argumentet om at det er sværere at sige nej til en person end en computer, kan vi også her argumentere for at det sikkert er nemmere at omgå sandheden og måske gå lidt hurtigt hen over spørgsmålene ved en computerbaseret undersøgelse fremfor undersøgelser med menneskelig interaktion.

En vigtig måling kan derfor være hvor lang tid respondenter bruger på besvarelsen. Med denne information kan man se manuelt på ekstremerne (personer der eksempelvis har besvaret 80 spørgsmål på 1½ minut) og således "luge ud" i eventuelt fejlbehæftede/tomme besvarelser. Som det ses nederst i Boks 3 kan man dog ikke forvente at dette vil udgøre en stor del af besvarelsene ved en internetbaseret undersøgelse. Vi anbefaler derfor at man gør sig tekniske overvejelser i forbindelse med at måle udfyldningstiden.

Internet som dataindsamlingsmedie

Sammenlignet med traditionelle dataindsamlingsmetoder stiller en internetundersøgelse os overfor nye og interessante muligheder, og naturligt nogle faldgruber. Vi vil i dette afsnit se på hvilke muligheder en internetbaseret dataindsamling giver os. Ydermere vil afsnittet indeholde gode råd baseret på erfaringer ved udvikling af spørgeskema til den Store Danske Internetundersøgelse 1999, [znail99]

Tekniske muligheder og begrænsninger

Ved at foretage dataindsamling via et internetbaseret spørgeskema har vi nogle helt oplagte fordele i form af elektronisk databehandling. Vigtigst er det at data på denne måde bliver indsamlet direkte til en database og kan derfor umiddelbart analyseres. Dette kan være temmeligt arbejdsbesparende både i forbindelse med indsamling, men også i forbindelse med de første faser i en dataanalyse hvor det gælder om at få overført de indsamlede data til en form hvor man kan foretage elektronisk databehandling.

Ydermere har vi den fordel i forbindelse med dataindsamling at vi kan fastholde respondentens svar fuldstændigt. Man kan kun besvare et givet spørgsmål på én måde og altså ikke sætte krydset midt mellem 2 felter som eksempelvis et papirbaseret spørgeskema.

Eksempel på dataindtastning

Vi vil her indledningsvis opsummere de konkrete muligheder for dataindsamling via internet. World wide web teknologien har fra starten haft muligheden for at lade brugere interagere med web-siden ved hjælp af form-elementer. Disse indtastningsmuligheder følger defacto standarderne for interaktion mellem menneske og computer - såsom at en række felter hvoraf kun ét kan vælges (også kaldet radiobuttons) fremstår som runde felter og en række felter hvor flere kan vælges fremstår som kantede felter.

Indsamles data via internet er man i den situation at man ikke kender brugerens opsætning nøjagtigt. Der kan være stor forskel på at opleve en internetside fra en Netscape version 2 på et UNIX system og til en Microsoft Explorer version 5 på et Windowsbaseret system.

Dette stiller store krav til udviklere af internetsider i forbindelse med kompatibilitet.

Internetteknologien udvikles konstant, nye HTML-standarder opstår og muligheder med JavaScript og Java eller Microsoft ActiveX giver helt nye muligheder for interaktion med brugeren. Men som nævnt ovenfor kender man

som regel ikke sine brugeres opsætning, hvilket styresystem de bruger, hvilken browser de bruger - og vil man derfor henvende sig til flest mulige internetbrugere kan dette kun opnåes ved at bruge de helt fundamentale muligheder HTML giver os.

Da størst mulig browserkompatibilitet er ønskeligt i de internetundersøgelser denne rapport omhandler, vil vi derfor ikke her gå i dybden med de muligheder der ligger ved fx at kombinere standard HTML med JavaScript eller ved at opbygge hele skemaet i fx Java.

Nedenstående eksempler på dataindtastning via internet er taget fra "*den store danske internetundersøgelse 99*". I tillægget til denne rapport ses iøvrigt en kopi af hele spørgeskemaet.

Standard tekstindtastning

Herunder ses to tekstfelter hvor respondenterne kan skrive tekst. Det er muligt at begrænse længden af tekststrengen i forbindelse med indtastning og i forbindelse med at formularen afsendes kan der på serversiden kontrolleres for at teksten overholder et bestemt format eller om feltet udfyldes eller ej. Såfremt brugeren laver en fejl i vore øjne kan vi bede ham taste om.



The image shows a screenshot of a web form. It contains two text input fields. The first field is labeled "Navn" and contains the text "Anders Blauenf". The second field is labeled "Email" and is empty. Below the "Email" field is a checkbox with the text "Deltagelse i lodtrækningen om en rejse".

Yderligere ses (under feltet med email) et enkelt afkrydsningsfelt - en checkboks der fungerer som en ja/nej indikation. Vi skal herunder se mere på hvad checkbokse kan bruges til.

Én svarmulighed - radiobuttons og lister

I gamle analoge bilradioer havde man ofte et system med forvalgte stationer. Var en station valgt (trykket ind) og man derefter valgte en ny, sprang den første knap ud. Dette har givet navn til et meget benyttet brugerinterface - radiobuttons.

Hvorledes betragter du dit eget niveau som internetbruger ?

- Helt nybegynder
- Ikke helt nybegynder (men stadig ofte i tvivl om basale ting)
- Almindelig bruger (Behersker de basale ting)
- Hvil bruger (Sjældent i tvivl)
- Superbruger (Aldrig i tvivl. Ved hvorledes web-sider er opbygget)

Respondenten kan kun vælge en af de fremstillede valgmuligheder, vælges en ny fravælges den forrige automatisk.

Er man i den situation at respondenten stadig kun skal vælge én ud af en meget lang liste - eksempelvis alle lande i verden, alle biograffilm der spiller i øjeblikket eller lignende vælger man som ofte at bruge et listfelt (også kaldet combobox). Som det ses her kan man på den måde presse flere felter ind på mindre plads. Således kan man gøre sit spørgeskema mere overskueligt.

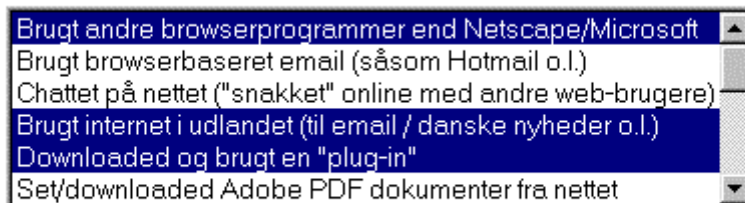
Arbejdsstatus	>>> vælges fra denne liste <<< ▼
Bopælsregion	>>> vælges fra denne liste <<< ▼
Bopælspostnummer	>>> vælges fra denne liste <<< ▼ landdistrikt landsby - under 1000 indb. by - under 30.000 indb. by - 30-100.000 indb. by - 100.000 - 250.000 indb. Århus / København bosat i udlandet

Flere svarmuligheder - checkbokse og lister.

Checkbokse bruges normalt i forbindelse med en liste hvor man gerne vil give brugeren mulighed for at afkrydse flere elementer. Det kan være listen af lande man har besøgt eller lignende. Herunder ses et eksempel på disse checkbokse hvor vi i forbindelse med at afdække erfaringen på nettet, spurgte vore respondenter om hvilke ting de havde prøvet.

- Brugt andre browserprogrammer end Netscape/Microsoft
- Brugt browserbaseret email (såsom Hotmail o.l.)
- Chattet på nettet ("snakket" online med andre web-brugere)
- Brugt internet i udlandet (til email / danske nyheder o.l.)
- Downloaded og brugt en "plug-in"
- Set/downloaded Adobe PDF dokumenter fra nettet
- Set web-tv (realvideo eller lignende)
- Downloaded programmer
- Klikket på en bannerreklame
- Læst i en nyhedsgruppe
- Skrevet i en nyhedsgruppe
- Abonneret på en mailingliste
- Handlet på internet med VISA-kort el lign
- Handlet på internet pr. efterkrav
- Opsøgt en internetcafé for at opnå internetforbindelse
- Udviklet internetsider

Igen findes der en variant til ovenstående interface. Hvis listen er meget lang - alle lande i verden, alle produkter i et produktkatalog eller lignende vil man overveje at benytte en list-visning i stedet. Med en sådan visning ville ovenstående indtastning se således ud:

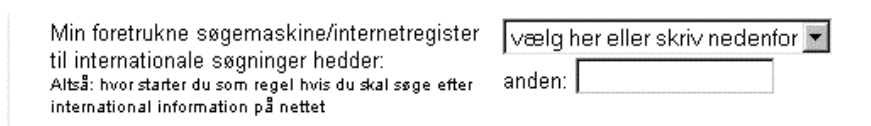


Altså igen samme information på mindre plads. Dog fremstår det en smule mere uoverskueligt, og sætter flere krav til respondenterne om fx at kombinere CTRL-tasten og SHIFT-tasten med musen for at vælge flere elementer på en gang.

Eksempler på kombinerede dataindtastningsfelter:

Vi vil herunder vise forskellige metoder at kombinere forskellige dataindtastningsfelter. Vi så allerede under tekstfelter at man kunne benytte en checkbox med én mulig værdi som et ja/nej-felt. Ved fx at kombinere en

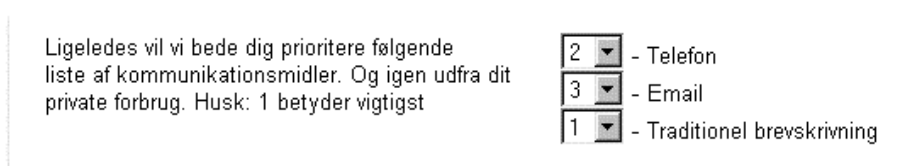
comboboks med et tekstfelt kan vi give brugeren mulighed for at vælge fra en liste og hvis listen ikke indeholder valget så kan det skrives manuelt:



Min foretrukne søgemaskine/internetregister til internationale søgninger hedder:

Altid: hvor starter du som regel hvis du skal søge efter international information på nettet

Et andet eksempel på at samle flere indtastningsfelter til at få oplysninger om én ting ses herunder hvor 3 combo-bokse giver respondenterne mulighed for at prioritere en liste af valgmuligheder. Dette giver yderligere mulighed for at tjekke om 2 af elementerne er ens - og i så fald eventuelt bede brugere taste om.



Ligeledes vil vi bede dig prioritere følgende liste af kommunikationsmidler. Og igen ud fra dit private forbrug. Husk: 1 betyder vigtigst

<input type="text" value="2"/>	- Telefon
<input type="text" value="3"/>	- Email
<input type="text" value="1"/>	- Traditionel brevskrivning

Man kan forestille sig adskillige andre kombinationsmuligheder ved at sammenholde forskellige indtastningsfelter - kun kreativiteten sætter grænsen.

Opsummering, dataindsamlingsfelter:

Vi har herover beskrevet de fundamentale muligheder for dataindsamling via internet. Vi har set på følgende felter og datatyper:

- Almindelig tekst. Udover det almindeligt indtastningsfelt vi har set, findes iøvrigt også et større felt til fx kommentarer - dette kan ses i bilaget.
- Liste hvor kun én skal vælges - her bruges ved få muligheder som regel radiobuttons og ved mange svarmuligheder bruges combo-bokse.
- Liste hvor mange kan vælges - her bruges ved få muligheder som regel checkbokse og ved mange svarmuligheder kan en åben scroll-liste benyttes.
- Yderligere har vi set hvordan man ved at kombinere ovenstående indtastningselementer kan opnå alternative muligheder for at stryge brugerens indtastning.

Konklusioner, erfaringer

Med baggrund i [znail99], og brugerkommentarer herved, er det vores erfaring at overdreven brug af datavalidering ikke kan anbefales. Hvis brugeren konstant gøres opmærksom på at indtastning ikke overholder det rette format, eller felter ikke må være tomme, irriteres respondenterne og det er ikke ønskeligt. Konsekvensen kan være at respondenterne giver op, eller endnu værre at

respondenten blot udfylder "det man tror man skal" for at komme videre - altså afgiver forkerte oplysninger. Det er i den forbindelse yderligere vores holdning at respondenten skal have mulighed for ikke at besvare enkelte spørgsmål - dette kan endda bidrage til analysen idet ej besvaret kan tolkes som "det ved jeg ikke noget om".

Ydermere kan vi anbefale at give respondenten frie rammer til selv at taste egne værdier ind samt kommentarer. 1265 respondenter (42%) har tilføjet kommentarer til spørgeskemaet ved 99-undersøgelsen. Selvom kommentarer ikke indgår direkte i data-analysen men udgives som særskilt bilag, er det vores fornemmelse at dataindtastning gøres behageligere ved at man kan komme med små meningsudbrud i tekstfelter.

En anden vigtig erfaring, er at vi kan anbefale opbygning af spørgeskemaet i selvstændige sider. Så snart man bevæger sig ud over 10-15 spørgsmål vil det blive alt for uoverskueligt for respondenten at blive præsenteret for det hele på én gang. Man kunne endda risikere at respondenter indtastede de første værdier (navn og email) for at deltage i konkurrencen, derefter besluttede at spørgeskemaet var alt for langt, og derfor fortsatte helt til bunden hvor de afleverede det - ikke ønskeligt. Denne sideopbygning må dog ikke gå udover browserkompatibiliteten, og skal således opbygges af teknologi på server-siden - ikke java/javascript eller andet der kører i browseren.

Afrunding, forslag til videre arbejde

Vi kan konkludere at internetbaseret dataindsamling stadig indeholder væsentlige bias-risici. I afsnittet omkring kommunikation med respondenter har vi argumenteret for forskellige svært analysér-bare og dermed svært korriger-bare biasmuligheder i forbindelse med at skabe og bibeholde kontakten til respondenter i en internetundersøgelse.

I afsnittet omkring den konkrete dataindsamling via internet, beskrives dog hvorledes internet stiller store muligheder i forbindelse med selve dataindsamling - både i form af fordele for respondenter og for dataindsamlere.

Kombination af RDD og internet

En naturlig konsekvens af ovenstående ville være at kombinere allerede kendte kontaktmetoder med den internetbaserede dataindsamling. Man kunne forestille forsøg med at skabe kontakt til respondenter via RDD (random digit dialing) og dernæst, med personlig kontakt, høre den potentielle respondenter om man ville udfylde et spørgeskema via internet. Vi kan dog argumentere for problemer ved denne metode, idet besvarelsen af internet-undersøgelser som oftest vil ske mens man alligevel er "på nettet". Fx kan vi se i [znail99] at ganske mange respondenter har besvaret skemaet i arbejdstiden - og dette kan ikke umiddelbart substitueres med ovenstående kombination med RDD og internet.

International dataindsamling

En vigtig pointe er at størstedelen arbejdsindsatsen ved at indsamle data via internet som regel ligger før undersøgelsen og stiger ikke væsentligt med antallet af respondenter. Således er det kun markedsføring af en undersøgelse, der bestemmer hvor mange respondenter det statistiske datamateriale skal bestå af. Således åbner internetbaserede undersøgelser nye muligheder for at udspørge meget store mængder af mennesker på en gang. En opgave der ville være meget krævende/umulig med de gængse dataindsamlingsmetoder. En oplagt videreførelse af internetundersøgelser ville eksempelvis være at foretage internationale sammenlignelige undersøgelser. Således kunne en undersøgelses-server i Danmark, over samme tidsperiode, modtage besvarelser fra hele verden med henblik på at sammenligne internetbrug i forskellige lande.

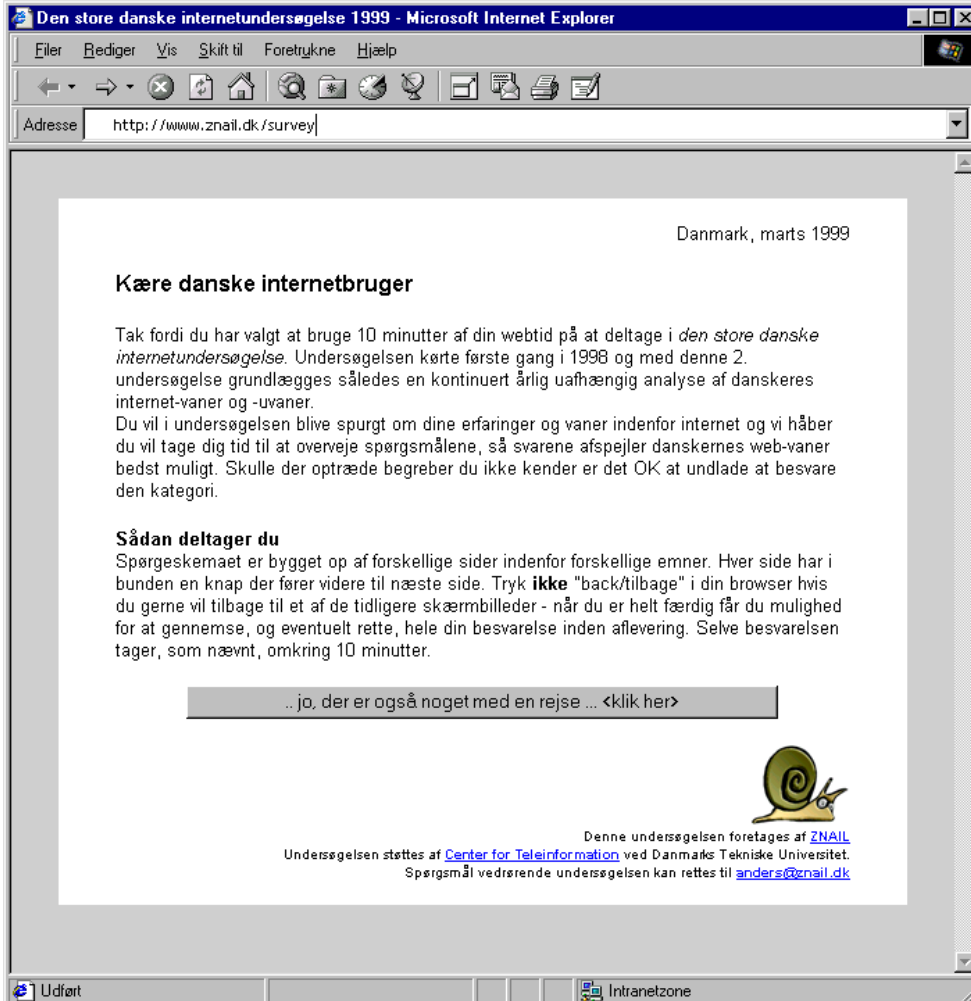
Begrebsforklaringer

Snakker man om internet opstår der tit begrebsforvirringer. Mange har forsøgt før os, men vi vil forsøge igen; en definition af de mest anvendte begreber omkring internet. Sprogforvirringen er tillige stor idet de fleste begreber stammer fra engelsk. Vi tager ikke her højde for sproget og bruger blot danske begreber hvor de findes og ellers engelske.

- Browser:** Et stykke software man benytter til at se internetsider og navigere imellem dem. Altså det program man surfer med - derfor kaldes det "at surfe på internet" også ofte "at browse".
- Cookie:** Mange internetsteder benytter sig efterhånden af cookies. En lille tekstfil gemmes på computeren og kan indeholde information om hvornår brugeren sidst besøgte internetstedet - eller hvorledes brugeren gerne ville have at forsiden på internetstedet skulle se ud.
- "Nettet":** En sproglig afart af Internettet. Bruges som regel til at betegne WWW-delen af internet. I princippet betegnelsen for forbindelsen mellem verdenens mange computere der udgør selve internettet. Dermed også betegnelse for alle andre aspekter ved internet; email, ftp, icq osv.
Begrebet at være "på nettet" kan dække over forskellige ting. Det bruges om at være online - have en forbindelse - og samtidig bruges det af andre om det at have muligheden for at blive online ("nu har jeg endelig købt et modem ... så er jeg også 'på nettet'").
- Surfe:** Bevæge sig rundt blandt websider.
- Website:** Et "logisk" område på internet - et "sted" - tilhørende og udarbejdet af samme firma/person/gruppe/organisation eller lignende. Kan bestå af mange websider. Behøver principielt set ikke ligge på samme server men gør det ofte.
- Webben:** se WWW
- Webside:** Én side på nettet - ofte ækvivalent med ét skærmbillede
- WWW:** World Wide Web. Den globale samling af websider. Opleves vha. en browser. Refereres også som "webben".

Bilag: selve spørgeskemaet

På de følgende sider ses en kopi af de enkelte skærbilleder, spørgeskemaet ved [znail99] var bygget op af.



The screenshot shows a Microsoft Internet Explorer browser window. The title bar reads "Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer". The address bar contains "http://www.znail.dk/survey". The main content area displays a survey introduction in Danish, dated "Danmark, marts 1999". The text explains the purpose of the survey and asks for user feedback. A button with the text "... jo, der er også noget med en rejse ... <klik her>" is visible. At the bottom, there is a logo of a snail and contact information for the Center for Teleinformation at the Technical University of Denmark.

Danmark, marts 1999

Kære danske internetbruger


Tak fordi du har valgt at bruge 10 minutter af din webtid på at deltage i *den store danske internetundersøgelse*. Undersøgelsen kørte første gang i 1998 og med denne 2. undersøgelse grundlægges således en kontinuert årlig uafhængig analyse af danskeres internet-vaner og -vaner.

Du vil i undersøgelsen blive spurgt om dine erfaringer og vaner indenfor internet og vi håber du vil tage dig tid til at overveje spørgsmålene, så svarene afspejler danskernes web-vaner bedst muligt. Skulle der optræde begreber du ikke kender er det OK at undlade at besvare den kategori.

Sådan deltager du

Spørgeskemaet er bygget op af forskellige sider indenfor forskellige emner. Hver side har i bunden en knap der fører videre til næste side. Tryk **ikke** "back/tilbage" i din browser hvis du gerne vil tilbage til et af de tidligere skærbilleder - når du er helt færdig får du mulighed for at gennemse, og eventuelt rette, hele din besvarelse inden aflevering. Selve besvarelsen tager, som nævnt, omkring 10 minutter.

... jo, der er også noget med en rejse ... <klik her>




Denne undersøgelse foretages af **ZNAIL**
Undersøgelsen støttes af [Center for Teleinformation](#) ved Danmarks Tekniske Universitet.
Spørgsmål vedrørende undersøgelsen kan rettes til anders@znail.dk

Udført Intranetzone

Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer

Adresse <http://www.znail.dk/survey>



A word from our sponsor ... vind en rejse !!

Vi har igen i år været så heldige at få en sponsoraftale. Så udover den store tilfredsstillelse der naturligvis ligger i at give sit besyv med i afdækningen af danskeres "net"-vaner [-:)], har du også muligheden for at deltage i lodtrækningen om et rejsegavekort. Et lille break fra den kolde danske vejr-udsigt kan man vel altid klare. Alle der deltager (seriøst) i undersøgelsen deltager automatisk i lodtrækningen om:

1 rejsegavekort på 5000 kroner sponsoreret af GateEleven
10 rejsegavekort på hver 500 kroner .. også sponsoreret af GateEleven
- såfremt du ikke ønsker at deltage i lodtrækningen kan dette fravælges på første side.

Undersøgelsen og dermed chancen for at vinde en rejse stopper mandag den 22. marts 1999, hvorefter der vil blive trukket lod blandt de deltagende. Vinderen får direkte besked og resultatet bliver offentliggjort via znail.dk. Man kan kun deltage i lodtrækningen med én besvarelse.

[Start spørgeskemaet <klik her>](#)

Intranetzone

Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer

Adresse http://www.znail.dk/survey

Først standardspørgsmålene - hvem er du i grunden ...
Vinderen bliver kontaktet via email - det er derfor vigtigt at dette felt bliver udfyldt korrekt.
Såfremt du ikke har en emailadresse kan du skrive din postadresse i samme felt.

Navn

Køn >>> vælges fra denne liste <<< ▼

Aldersgruppe >>> vælges fra denne liste <<< ▼

Email
 Deltagelse i lodtrækningen om en rejse
 OK at vi sender en mail ved undersøgelser fremover (altså højst et par mails om året ... læs isvrigt vores emailpolitik [her](#))

Arbejdsstatus >>> vælges fra denne liste <<< ▼

Bopælsregion >>> vælges fra denne liste <<< ▼

Bopælspostnummer

videre .. videre <klik her>

Udført Intranetzone

Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer

Filet Rediger Vis Skift til Foretrykne Hjælp

Adresse http://www.znail.dk/survey

Dernæst lidt om din intererfaring
Dette afsnit tjener til at få et overblik over hvor lang tid du har benyttet dig af "nettet" og hvor godt du behersker dette medie.

Hvilket år brugte du første gang World Wide Web >>> vælges fra denne liste <<<

- World Wide Web (www) er i denne forbindelse ensbetydende med at bruge en browser. Altså at surfe.

Hvilket år sendte/modtog du første gang internet-email ? >>> vælges fra denne liste <<<

- her menes en elektronisk postbrev sendt via internettet med adresseformatet navn@sted.land

Hvorledes betragter du dit eget niveau som internetbruger ?

- Helt nybegynder
- Ikke helt nybegynder (men stadig ofte i tvivl om basale ting)
- Almindelig bruger (Behersker de basale ting)
- Hvil bruger (Sjældent i tvivl)
- Superbruger (Aldrig i tvivl. Ved hvorledes web-sider er opbygget)

Hvor meget tid bruger du på nettet ? >>> vælges fra denne liste <<<

- her menes ikke den tid hvor du "er" på nettet, men laver noget andet. Kun den tid du aktivt bruger på at skrive mails, bruge www og så videre ...

Hvorfra har du adgang til nettet ?

- via arbejde
- via studie
- hjemmefra

andet:

Hvor mange emailadresser har du ? >>> vælges fra denne liste <<<

Hvor ofte tjekker du din email ? >>> vælges fra denne liste <<<

mere om din net-erfaring, samt en lille quiz <klik her>

Udført Intranetzone

Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer

Filer Rediger Vis Skift til Foretrykne Hjælp

Adresse http://www.znail.dk/survey

Mere omkring interneterfaring

Herunder ses en (ganske usorteret) liste af forskellige elementer ved at bruge internet. Kryds af ud for de ting du har prøvet ?

- Brugt andre browserprogrammer end Netscape/Microsoft
- Brugt browserbaseret email (såsom Hotmail o.l.)
- Chattet på nettet ("snakket" online med andre web-brugere)
- Brugt internet i udlandet (til email / danske nyheder o.l.)
- Downloaded og brugt en "plug-in"
- Set/downloaded Adobe PDF dokumenter fra nettet
- Set web-tv (realvideo eller lignende)
- Downloaded programmer
- Klikket på en bannerreklame
- Læst i en nyhedsgruppe
- Skrevet i en nyhedsgruppe
- Abonneret på en mailingliste
- Handlet på internet med VISA-kort el lign
- Handlet på internet pr. efterkrav
- Opsøgt en internetcafé for at opnå internetforbindelse
- Udviklet internetsider

Og så tager vi lige en lille quiz :-)

Herunder ses en række spørgsmål om internet generelt. Det gælder ikke om at svare mest rigtigt - det er blot for at undersøge din viden (og dermed danskernes generelle viden) om internetrelaterede emner. Svar derfor blot så godt du kan ...

Hvilken af følgende ting mener du er **absolut** nødvendige for at man på en computer kan bruge internettet ?

- har ingen anelse
- en browser
- en emailadresse
- en mus
- et modem
- et netværkskort
- netværksprotokollen TCP/IP
- 32 Mb Ram

Hvornår tror du selve konceptet bag internet blev udviklet - altså det her med at sætte computere sammen i et globalt netværk. >>> vælges fra denne liste <<<

Hvornår tror du world wide web blev udviklet - altså det her med at man kan klikke rundt imellem forskellige websider og finde information. >>> vælges fra denne liste <<<

ikke mere quiz .. videre ... <klik her>

Udført Intranetzone

Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer

Adresse http://www.znail.dk/survey

Hvad bruges "nettet" til ... og hvordan
 Hvad bruger du nettet til. Hvilke informationer finder du på "nettet" - og hvordan. Overvej hvorledes følgende udsagn passer på dig.

Mit primære brug af email kan bedst beskrives som
 arbejds/studie-relateret privat relateret lige dele arbejde/studie og privat

Mit primære brug af world wide web kan bedst beskrives som
 arbejds/studie-relateret privat relateret lige dele arbejde/studie og privat

Jeg finder dagsaktuelle nyheder, tv-programmer, vejrudsigt mm på "nettet"
 dagligt ugentligt sjældent aldrig

Jeg bruger en world wide web til at finde information af leksikal karakter
 dagligt ugentligt sjældent aldrig

Jeg bruger internettet når jeg vil underholdes / slappe af
 dagligt ugentligt sjældent aldrig

Hvis jeg skal søge job vil jeg bruge en jobdatabase på internettet
 ja nej ved ikke

Lidt mere om chat på nettet:
 - sæt kryds ud for de sætninger der passer for dig
 - jeg bruger chat til at hyggesnakke med folk jeg ikke kender
 - jeg har lært nye mennesker at kende via chat
 - jeg bruger chat til at hyggesnakke med folk jeg kender (som en erstatning/supplement til tlf.)
 - jeg bruger chat til at søge rådgivning
 - jeg bruger chat til at udveksle erfaringer indenfor et specifikt område
 - jeg udgiver mig nogen gange for en anden end jeg er når jeg chatter
 - jeg har aldrig brugt chat på nettet

... og lidt mere om søgning på nettet:
 Min foretrukne søgemaskine/internetregister til danske søgninger hedder:
 Altså: hvor starter du som regel hvis du skal søge efter dansk information på nettet anden:

Min foretrukne søgemaskine/internetregister til internationale søgninger hedder:
 Altså: hvor starter du som regel hvis du skal søge efter international information på nettet anden:

Hvad glemte vi ? Hvad bruger du derudover "nettet" til ? Hvorfor kan du ikke undvære det ? Hvornår vil du helst undvære det ?

Videre til irritationsmomenter på nettet ... <klik her>

Udført Intranetzone

Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer

File Rediger Vis Skift til Foretrykne Hjælp

Adresse http://www.znail.dk/survey

Irritationsmomenter

Hvad irriterer dig mest når du browser/surfer. Er det hastigheden, er det uoverskueligheden, søgefunktionerne eller andet. Vi bruger følgende skala: **1:** går aldrig én på, **2:** er lidt irriterende, og **3:** er drønirriterende.

World Wide Wait - at det hele går for langsomt. 1 2 3 ved ikke

Krav om plug-ins for at man kan opleve websites rigtigt: 1 2 3 ved ikke

For dyrt - telefonregningen bliver for stor. 1 2 3 ved ikke

Der er for meget grafik på websider. Det gør dem for langsomme. 1 2 3 ved ikke

Dårligt opdaterede web-sites - informationerne er for gamle 1 2 3 ved ikke

Browsers forskelligheder - Netscape og Microsoft Explorer opfører sig forskelligt: 1 2 3 ved ikke

Reklamer: 1 2 3 ved ikke

Dårlige søgeresultater - du føler ikke at du behersker søgemaskinerne optimalt: 1 2 3 ved ikke

Dårlige søgeresultater - du synes at søgemaskiner generelt er for dårlige: 1 2 3 ved ikke

Andre ting der irriterer når man surfer:

her har du muligheden for at komme ud med det, hvis du sidder med indestængt internet-irritation ... :)

Nå ... er vi ikke snart færdige ... <klik her>

Udført Intranetzone

Den store danske internetundersøgelse 1999 - Microsoft Internet Explorer

Adresse http://www.znail.dk/survey

Afrunding

Til sidst nogle mere brede spørgsmål omkring medieforbrug, dit syn på fremtiden, samt nederst et par spørgsmål angående din deltagelse i denne undersøgelse.

Først vil vi bede dig prøve at prioritere følgende liste af medier ud fra dit private forbrug. Det du mener er vigtigst for dig (og altså **nødigst** vil undvære) giver du værdien 1, og så fremdeles.

- Radio
 - TV
 - Internet
 - Aviser, blade

Ligeledes vil vi bede dig prioritere følgende liste af kommunikationsmidler. Og igen ud fra dit private forbrug. Husk: 1 betyder vigtigst

- Telefon
 - Email
 - Traditionel brevskrivning

Derefter bedes du overveje disse udsagn om fremtiden og krydse af ud for dem der passer på dig.

Jeg tror at jeg vil bruge væsentligt mere tid "på nettet" om 5 år end jeg gør i dag
 Jeg tror email bliver ligeså udbredt som telefon er i dag (altså at stort set alle har en emailadresse)
 Jeg tror at internettet helt kan udrydde andre medier (såsom aviser - blade eller TV)
 Jeg tror at størstedelen af mine dagligdags indkøb (mad - husholdning) vil ske via internettet indenfor 5 år
 Jeg tror at størstedelen af mine store indkøb (computere - cykler - møbler mv) vil ske via internettet indenfor 5 år
 Jeg tror internettet er kommet for at blive [bare rolig - dette sidste spørgsmål er blot for at se om du stadig er i live ... :-)]

Lidt om denne undersøgelse

Deltog du i denne undersøgelse sidste år ? ja nej

Hvor har du hørt om denne undersøgelse ?

Hvor tit deltager du i konkurrencer på nettet ?

Hvad synes du om længden af denne undersøgelse ?

For kort - kunne sagtens blive ved.
 Passende - god længde.
 For lang - min koncentration daledede.
 Alt for lang - jeg kunne slet ikke bevare koncentrationen.

det var så det... <klik her>

Udført Intranetzone