

Kapittel 3

Datainnsamling
Dataproduksjon

Produsere/samle data

- Kap 1: Utforske og analysere **gitte** data for en variabel ved hjelp av grafer og tall
- Kap 2: Analysere sammenhenger mellom **gitte** data for to variable
- Nå: Strategier for å produsere/samle inn data
 - Produsere data av god kvalitet
 - Vurdere kvalitet på data hentet inn av andre
 - Åpner dør til formell statistisk inferens, som besvarer spesifikke spørsmål med et gitt konfidensnivå

Typiske spørsmål

- Hvilke individer skal jeg studere (umulig å samle inn data på alle)?
- Hvilke variable skal jeg måle?
- Hvor mange individer?
- Hvordan velge ut individer?
- Hvordan dele opp i grupper (hvis sammenlikning av behandlinger)?
- +++

Feil: Anekdotisk bevis

- Trekker konklusjoner basert på utvalgte enkeltsituasjoner som vi hører om fordi de er spesielle i en eller annen sammenheng
- Behøver ikke å være **representative** for virkeligheten
- Eksempel
 - 'Man får kreft av stråling fra mobilmaster, fire kvinner i samme borettslag har fått kreft det siste halvåret...'

Statistisk bevis: data og analyse

- Tilgjengelige datasett
www, Statistisk sentralbyrå, Meteorologisk institutt, Helseregistre (unike i Norge), etc.
- Utvalgsundersøkelser
- Eksperimenter

Utvalg

- Typisk ved meningsmålinger og kartlegginger
 - General Society Survey (GSS): Intervjuer 3000 individer (i Norge typisk 1000)
 - **Utvalg** av en større **populasjon**
 - Studerer del for å samle informasjon om hele populasjonen
 - Må være **representativ**
 - Observasjonsstudie!

Observasjoner vs eksperiment

En **observasjonsstudie** observerer individer og måler variable av interesse, men prøver ikke å influere responsen (3.2)

Et **eksperiment** vil bevisst tillegge ulike behandlinger på individer for å observere responsen (3.1)

3.1 Design av eksperimenter

(Forsøksplanlegging)

Individer = **eksperimentelle enheter**

Når individene er mennesker kalles de ofte **personer**

Eksperimentell tilstand på enhet = **behandling**

Forklaringsvariabel ofte kalt **faktor**

Ulike **nivåer** av faktorer

Behandling resultat av kombinasjoner av nivåer av flere faktorer

En faktor: Klassestørrelser

- Effekt på - karakterer? - tilstedeværelse på skole? -videre utdanning?
- Eksperiment: 6385 barn (**individer/personer**), 3 **behandlinger**, dvs **faktoren** klasstype har 3 **nivåer**
 - Vanlig klasse (22-25 barn), en lærer
 - Som over men med assistent
 - Liten klasse (13-17 barn)

Alt annet likt

- Små klasser: bedre karakterer senere, flere besto, flere gikk videre, spesielt gunstig for minoritetselever
(Eks. 3.7 s.166)

Flere faktorer: Effekt av reklame

- Reklameinnslag på TV – Eks. 3.8 s. 167
- Effekt avhengig av -hvor lenge (?), -hvor ofte (?)
- Individier: Lavere grads studenter
- Alle 40 min TV-program inkl. reklame for digitalt kamera
- 2 nivåer av faktoren **lengde**: 30 sek eller 90 sek reklame
- 3 nivåer av faktoren **antall repetisjoner**: 1, 3 eller 5 repetisjoner
- Respons: Holdning til kamera (spørreskjema)
- To faktorer, henholdsvis 2 og 3 nivåer -> Totalt 6 ulike behandlinger (kombinasjoner av nivåene til de to faktorene)

Kan studere kombinerte effekter av ulike faktorer, som ikke er mulig ved bare å se på en faktor alene

Effekt av reklame: 6 ulike behandlinger

		Factor B Repetitions		
		1 time	3 times	5 times
Factor A Length	30 seconds	1	2	3
	90 seconds	4	5	6

Figure 3-2
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

Sammenliknende eksperimenter

Typisk for laboratorie-eksperimenter:

Enkle design

En enkel behandling

Behandling -----> **Observert respons**

Bør **sammenligne** to grupper, der kun den ene får behandlingen

Magefrysning

Effekt av behandling for magesår

Studie i Journal of the American Medical Association:

Magefrysning (reduserer syre) ---> Observert smertelindring

Placebo-effekt?

Mange responderer godt på en behandling bare de *tror* den virker

Senere studie: To grupper pasienter, en med samme behandling, en med tilsvarende behandling men ingen frysing

Viste ingen forbedring ved frysning

Bruk av kontrollgrupper

- Magefrysning: Sammenblandede variable (placebo underliggende variabel)
- Kan unngås ved bruk av **kontrollgruppe** (spesielt i medisin)
- Kontrollgruppe: Ingen behandling eller placebo
- **Skjevhet** («bias»): Systematisk favorisering av visse utfall
- Kontrollgrupper gjør oss i stand til å kontrollere effekter av underliggende variabler, ved at eneste forskjellen mellom behandlet gruppe og kontrollgruppen er behandlingen
- Kontrollert studie: viktig prinsipp for å unngå skjevhet/bias

Hvordan allokere individer til grupper?

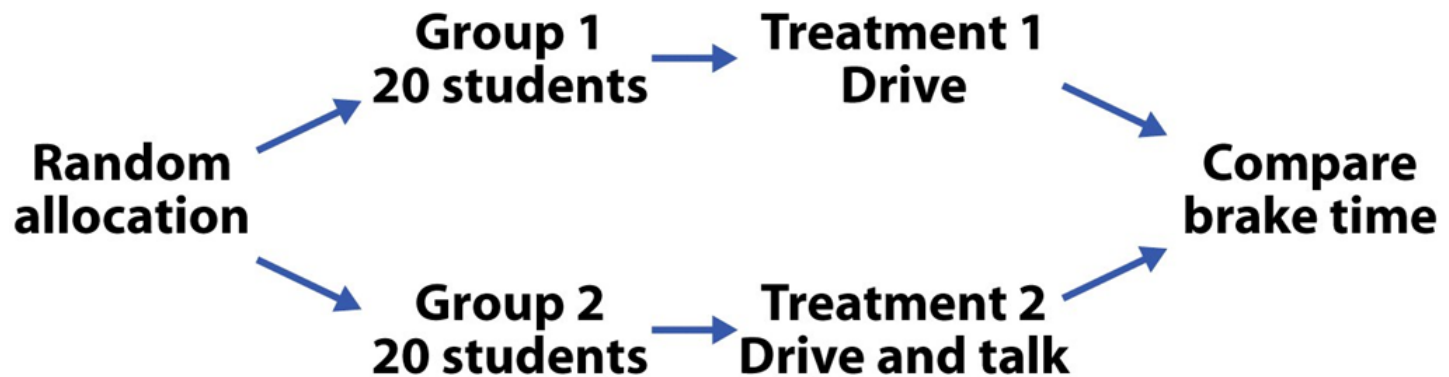


Figure 3-3
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

Randomisering

- Viktig at grupper som sammenlignes har like egenskaper
 - likt jordsmonn for sammenligning av korn
 - likt sykdomsforløp i behandlingsgrupper
- Systematiske forskjeller ---> skjevhet
- Mulighet: «matche» ulike egenskaper (alder, kjønn, helsetilstand etc) for å få like egenskaper i gruppene
- Vanskelig, kan være faktorer en ikke vet om
- Strategi: **Randomisering**, alloker individer tilfeldig til grupper. Unngår **systematiske** forskjeller

Randomiserte sammenliknende eksperimenter

- **Randomisering** produserer to grupper som vi forventer er like
- **Sammmenlignende design** hjelper til å sikre mot påvirkning av andre faktorer
- Tilfeldigheter og naturlig variabilitet fremdeles tilstede. Kan reduseres ved **replikasjoner** (mange individer)

Statistisk signifikans

- Forskjeller kan fremdeles skyldes **tilfeldig variabilitet**
- Forskjeller så store at det er usannsynlig at det kun skyldes tilfeldigheter kalles **statistisk signifikante**
- Alt vi skal gjennomgå etter midtveiseksamen (kap. 6,7,10,11) handler om dette!

Hvordan randomisere

- Nummererer individene og bruker software til å allokere tilfeldig
- Tabell av tilfeldige tall (utdatert?)

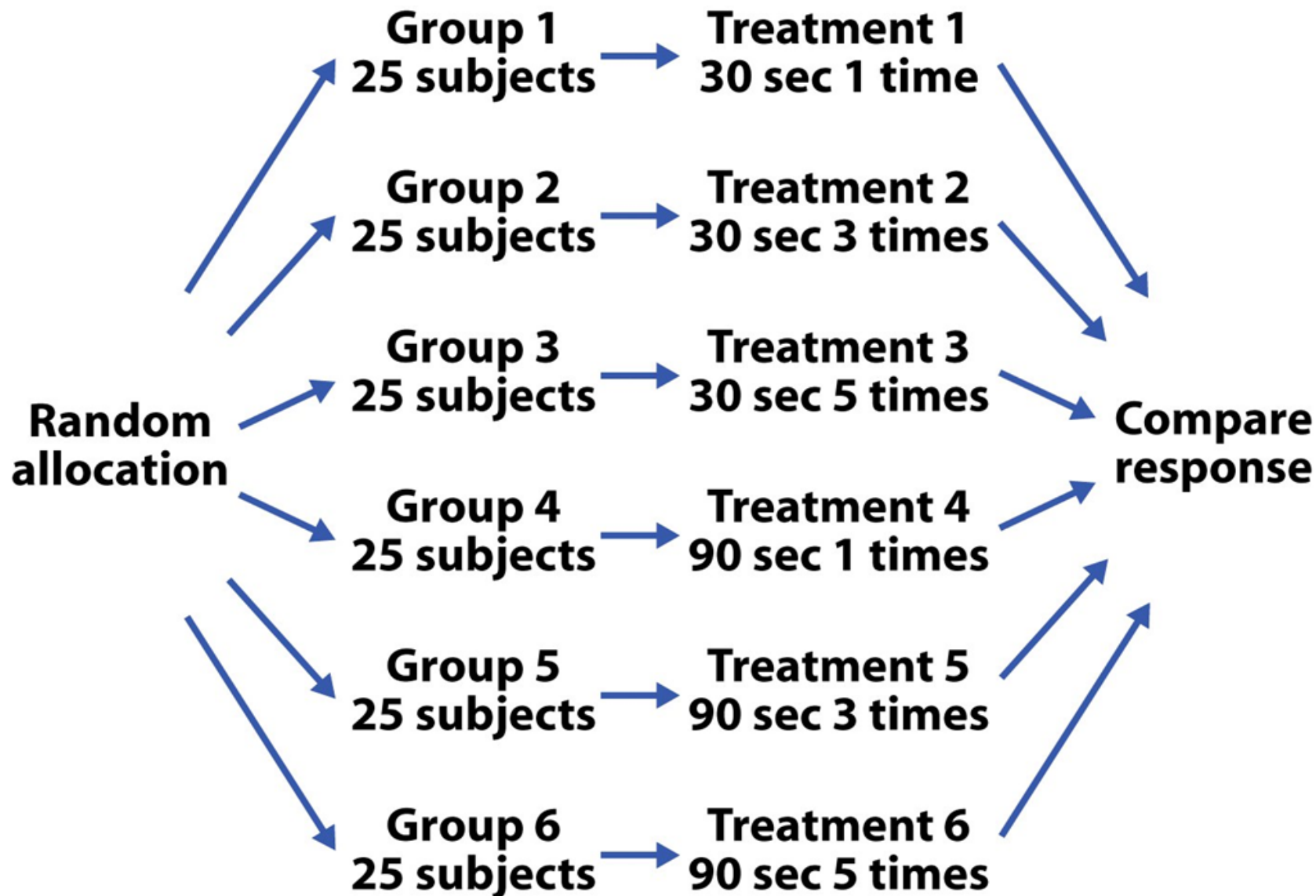


Figure 3-4
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

OBS! for randomiserte sammenlignende eksperimenter

- Forutsetter at alt utenom behandling er likt
- Magefrysing: Behandlings- og kontrollgruppe ble gitt (nesten) identisk behandling
- Studien var også **dobbelt blind**: Hverken pasient eller personell visste hvilken behandling som blir gitt
 - Hindrer at f.eks. en lege trekker en konklusjon om ingen effekt fordi han vet at pasienten fikk placebo

- Eks. 3.14 s. 176 Effekt av marijuana:
Kontrollgruppe gitt sigaretter istedet, men disse kjente forskjell (ikke-blind studie)
- Eksperimenter kan ikke alltid generaliseres.
Genetisk innflytelse på oppførsel: Et gen «slås av» i mus, sammenliknes med kontrollgruppe
 - Ulike studier gir ofte ulike resultater
 - Nye studier med genetiske like mus ga også forskjeller
 - Ulike laboratorie-fasiliteter ga store forskjeller

Parvise design

- Randomisert design enkelt, men mulig med forbedringer
- Parvise design: Individu settes sammen i par
- Par likere enn individer fra ulike par
 - Tvillingstudier
 - Mus fra samme kull
 - Samme individ får begge behandlinger (er sin egen kontroll)
 - Randomisert i hvilken rekkefølge de fikk behandling
- Randomisering innen par!

Blokk design

- Blokk: Gruppe av individer som er like før eksperiment
- Fordeling/randomisering på behandling skjer innen hver blokk
- Kontrollerer effekten av utenforliggende variable ved å bringe dem inn i eksperimentet som blokker
- Gir oss muligheten til å separere konklusjonene for hver blokk
- Gir mere presise konklusjoner fordi den systematiske forskjellen mellom blokkene fjernes fra analysen av effekten av behandlingene
- Eksempel: Blokkdeling etter kjønn i en studie av effekten av ulike behandlinger for kreft

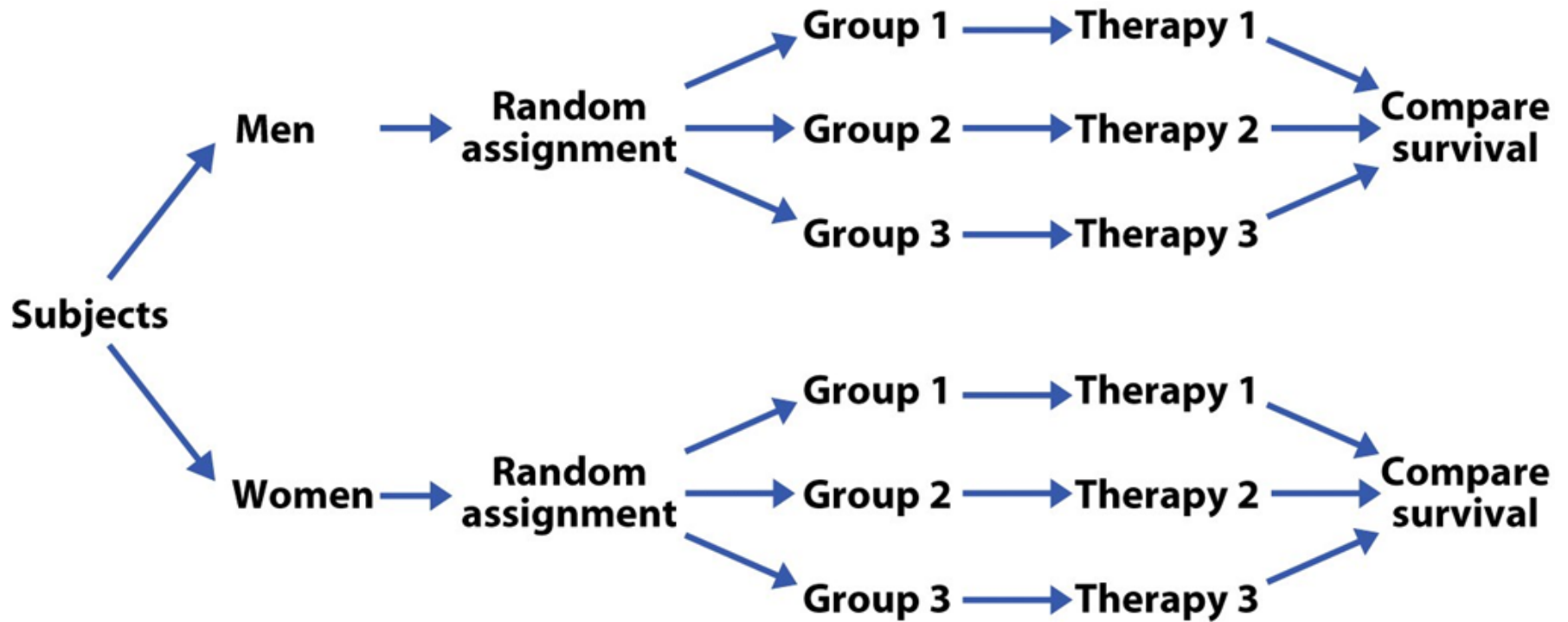


Figure 3-5
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

- Kap 3.1: Produsering av data:
 - Design av styrte eksperimenter
 - Sammenlignende design/kontrollgrupper
 - Parvis/blokk design

- Kap 3.2: Produsering av data:
 - Design av observasjonsstudier
 - Randomiserte utvalg
 - Stratifisering
 - Replikasjoner

Observasjonsstudie vs eksperiment (igjen)

- Et **eksperiment** vil bevisst tillegge ulike behandlinger på individer for å observere respons på behandling (3.1)
- En **observasjonsstudie** observerer individer og måler variable av interesse uten å tillegge en behandling, dvs prøver ikke å influere responsen (3.2)
 - Må finne et passende **utvalg** å observere

Design av utvalg

- Andel studenter som er negative til privatisering?
- Andel ungdom som ser tv-reklame om ny sportssykkel?
- Meningsmåling før kommunevalget i Oslo
- Gjennomsnittlig inntekt i Skandinavia?
- Ønsker informasjon om en stor **populasjon**, f.eks. alle universitets-studenter, alle ungdom mellom 13-18, hele Oslos myndige befolkning, ...
- Tid, kostnad, etc -> Kan bare undersøke et **utvalg** av populasjonen
- Design av utvalg: Hvordan velge ut utvalg

Dårlige utvalg

- Fabrikk produserer ruller av tynt stål
 - Ønsker å sjekke kvalitet
 - Ber tekniker kutte små biter til undersøkelse
 - Tekniker velger kun «gode biter»
- Debattprogram og sms-målinger
 - Ønsker meningsmåling i populasjon
 - Teller opp andel for/mot fra sms-innsendinger
 - Ikke representativt for populasjonen

Frivillig respons-utvalg

- Et **frivillig respons-utvalg** består av personer som velger å svare på en generell forespørsel
- Slike utvalg er **forventningsskjeve** fordi personer med sterke meninger, spesielt negative meninger, svarer oftere

[forventningsskjev=skjev=biased]

Unngå forventningsskjevhet

- Tilfeldig, objektivt valg av individer til utvalget gir et representativt utvalg av populasjonen
- Forventningsskjevhet unngås
 - Unngår skjev utvelgelse (favorisering) fra den som gjør utvalget (som med stålbitene)
 - Unngår skjev utvelgelse fra frivillige respondenter

Enkelt tilfeldig utvalg

- n individer trukket ut av total populasjon slik at alle individer har like stor sjanse for å komme med.
- Kalles enkelt tilfeldig utvalg (SRS: simple random sample) av størrelse n
- SRS er byggesten for mer komplekse metoder for utvalg

Stratifisert tilfeldig utvalg

- Del populasjon inn i grupper av lignende individer, kalt **strata**
- Velg et enkelt tilfeldig utvalg innen hver strata
- Kan gi mer eksakt informasjon enn ett enkelt tilfeldig utvalg
- **Strata** for utvalg er tilsvarende til **blokker** i eksperimenter

Eksempel stratifisering

- Meningsmålinger
- Strata: Fylker
- Trekker ett enkelt tilfeldig utvalg fra hvert fylke
- Reduserer variabilitet som skyldes forskjeller mellom fylker

Underdekning og mangel på respons

- Tilfeldige utvalg fjerner skjevheter
- Design av utvalg for store populasjoner krever gode design
- **Underdekning**: Noen grupper glemt (f.eks. telefonintervjuer utelater de husholdningene som ikke har telefon)

Ikke-respons (individer ikke tilgjengelige eller nekter å respondere):

Større manglende respons i urbane strøk enn på landsbygda, kan gi skjevheter dersom det ikke tas hensyn til

Dersom man erstatter de som ikke responderer med andre som ikke ligner på de som responderer, kan det også gi skjevheter

Andel ikke svarende

- Eks. 3.25 s. 191. Current Population Survey (utført av myndighetene): Lav andel ikke-respons
 - 4% nekter å svare, 3-4% kan ikke nås
- General Social Survey (utført av et universitet): Høyere andel ikke-respons (30%)
- Problem: De som ikke svarer kan ha andre meninger/egenskaper enn de som svarer, noe som medfører forventningsskjevhet

Spørreskjema/intervju: Respons-skjevhet

- Kan få forventningsskjevhet pga. egenskaper til den som intervjuer
 - Intervjuer til levekårsundersøkelse utført av representant fra myndighetene -> respondenter lyver om ulovlig oppførsel
- Spørsmålsform kan også medføre forventningsskjevhet
 - Ledende
 - Vanskelige å tolke/misforståelser

Annet eksempel på skjevt utvalg: Biologi

- Av interesse: Vekt av lemmen
- Utvalg: Lemmen tatt i feller
- Har lemmen som ikke har gått i feller samme vekt som de som har gått i feller?

3.3 Statistisk inferens intro

- Hvordan overføre informasjon fra utvalg til informasjon om populasjon?
- Eksempel: Intervju av et utvalg på 2500
 - 1650 (66%) finner kjøp av klær frustrerende og tidkrevende
 - Hva er sannheten om hele populasjonen (220 millioner USA)?
- **Statistisk inferens**: Trekke konklusjoner om større populasjon fra det mindre utvalget

Statistisk inferens

- **Statistisk inferens:** Trekke konklusjoner om større populasjon fra det mindre utvalget
- For å gjøre det, må vi skille mellom
 - tall som beskriver et utvalg (**observatorer**)
 - tall som beskriver populasjonen (**parametre**)

Parametre og observatorer

- **Parameter**
 - Et tall som beskriver populasjonen
 - Fast tall, men ukjent
- **Observator**
 - Tall beskrevet av utvalg
 - Kjent, men endres fra utvalg til utvalg
- Bruker observator for å **estimere** (anslå) parameter

Parametre og observatorer

- Eksempel shopping: Intervju av et utvalg på 2500
 - 1650 (66%) finner kjøp av klær frustrerende og tidkrevende
 - Hva er sannheten om hele populasjonen (220 millioner)?
 - p : **parameter** som beskriver andel av **populasjon** som finner kjøp av klær frustrerende og tidkrevende
 - $\hat{p} = 1650/2500=0.66$ er **observatoren** for **utvalget**, og kan brukes som et **estimat** av p

Utvalgsvariabilitet

- Nytt utvalg med nye 2500 personer vil gi nytt estimat \hat{p} , antageligvis forskjellig fra 0.66
- **Utvalgsvariabilitet**: Verdien av en observator varierer når man repeterer tilfeldige utvalg
- Tilfeldig utvalg
 - Fjerner skjevhet
 - Vil alltid ha variabilitet
- Men: Variasjonen følger et **forutsigbart mønster**
 - Statistisk inferens baserer seg på å spørre seg hva som skjer hvis prosedyren blir **repetert** mange ganger. Det sier noe om hvor **pålitelig** en prosedyre er

Hvordan måle utvalgsvariabilitet?

- Ta mange utvalg fra samme populasjon
- Beregn \hat{p} for hvert utvalg
- Lag et histogram av \hat{p}
- Undersøk histogrammet for form, senter og spredning til fordelingen
- I praksis
 - For dyrt/tidkrevende å se på mange utvalg
 - Alternativ: **Simuler** med datamaskin (*imitererer* mange utvalg)

Utvalgsvariabilitet: Simulering av utvalg

- Later som om vi vet at andelen i populasjonen som synes kjøp av klær er frustrerende og tidkrevende er 60%, dvs $p=0.6$
- Trekker 1000 utvalg av størrelse $n=100$ fra en slik populasjon (trekker 100 personer 1000 ganger fra en populasjon med $p=0.6$) *Hvordan?*

For hvert av de 1000 utvalgene beregnes \hat{p} ,
lager så et histogram over de 1000 verdiene av \hat{p}

Hvordan endres variabiliteten i forhold til
størrelsen n på utvalget? Gjenta simulering for
utvalg av størrelse $n=2500$

Figurer:

Tilfeldig utvalg på størrelse 100

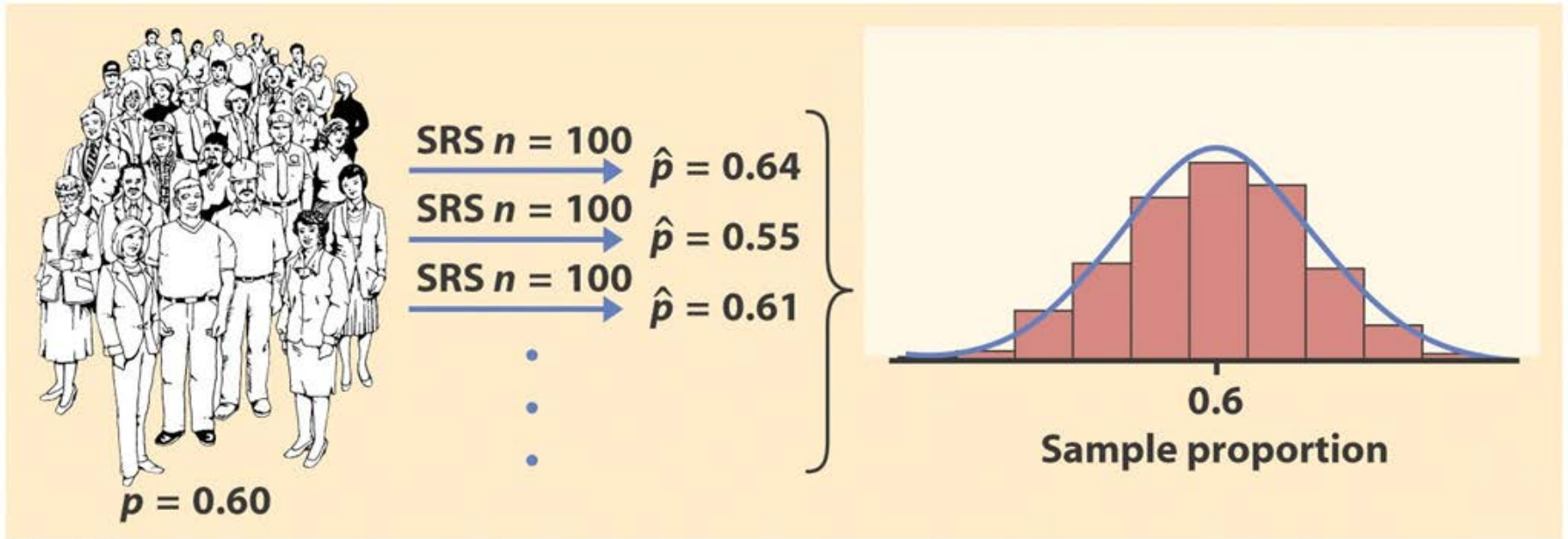


Figure 3-9
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

Tilfeldig utvalg på størrelse 2500

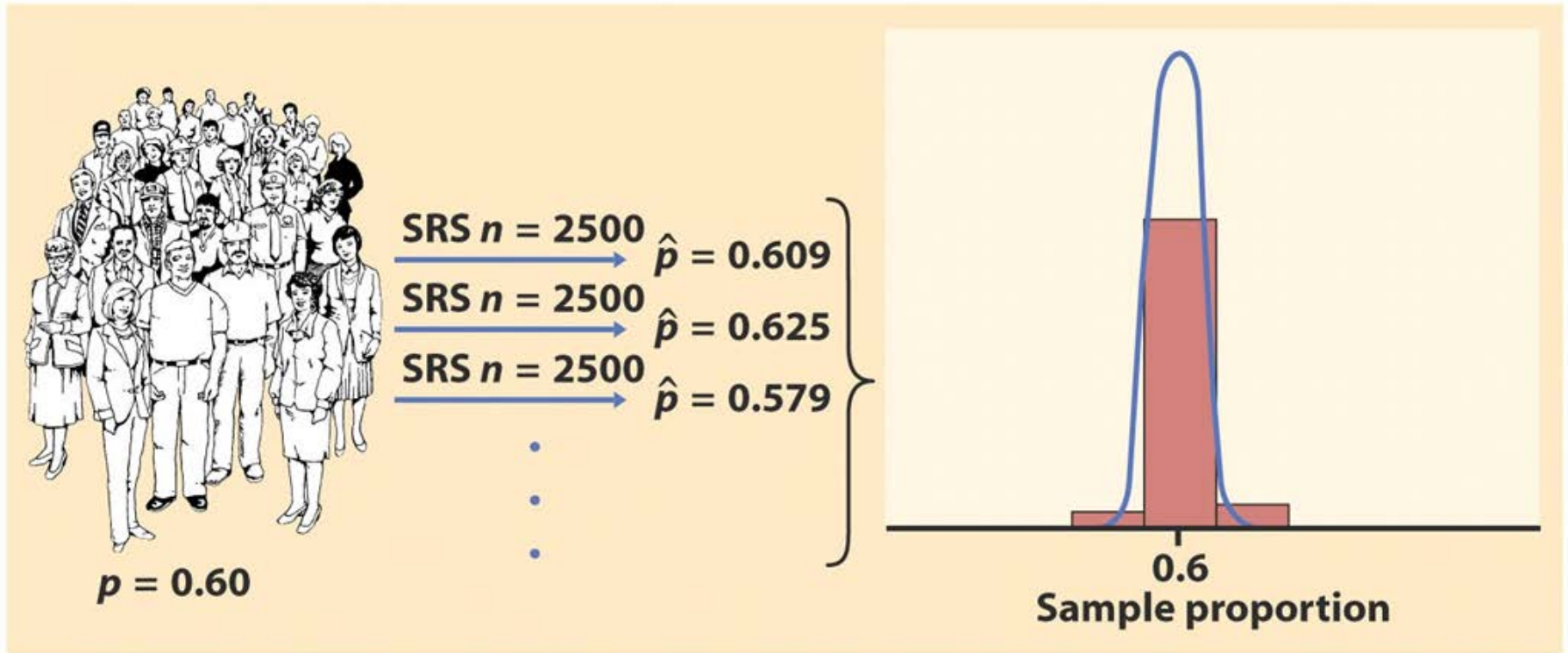


Figure 3-10
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W.H. Freeman and Company

- Histogrammene gir oss en ide om hvordan utvalgsvariabiliteten avhenger av n
- Gir et bilde av **utvalgsfordelingen** til observatoren \hat{p} for to forskjellige verdier av utvalgsstørrelsen n

Utvalgsfordeling

- **Utvalgsfordelingen** for en observator er fordelingen av verdier tatt over alle mulige utvalg av samme størrelse n fra populasjonen
- Simulering: Tilnærming av sann fordeling (eksempel: vi så kun på 1000 utvalg, *ikke alle mulige*)
- Sannsynlighetsteori: Kan noen ganger gi eksakt fordeling – dette kommer senere
- Beskriver fordeling ved form, senter og spredning

- Form: Begge histogrammene ser ut til å stemme bra med normalfordelingen (kvantilplott bekrefter dette (for $n=100$))
- Senter: Begge histogrammer er sentrert i 0.6, ingen tendens til at verdiene er systematisk høyere eller lavere enn 0.6, dvs \hat{p} har ingen forventningsskjevhet som estimator for p
- Spredning: Mye mindre spredning for $n=2500$ enn for $n=100$

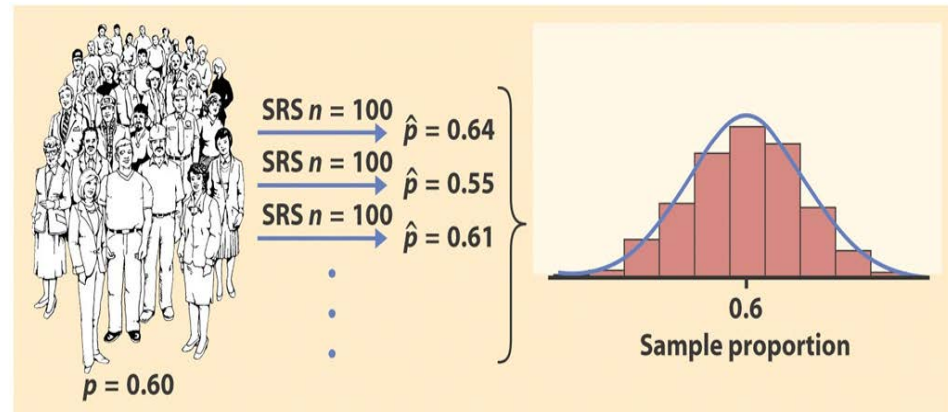


Figure 3-9
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

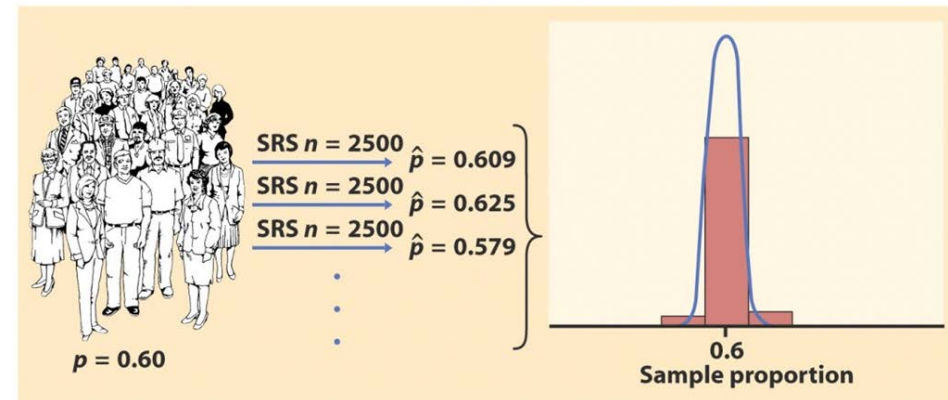


Figure 3-10
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

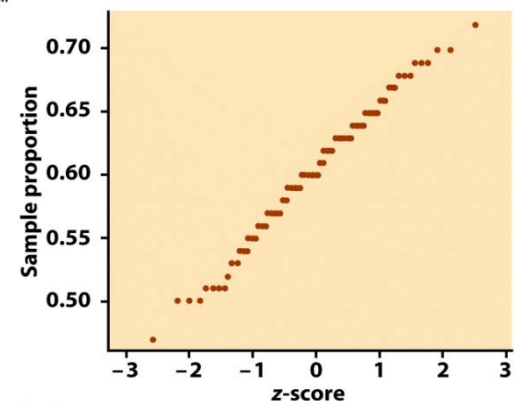


Figure 3-11
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

Forventningskjevhet og variabilitet

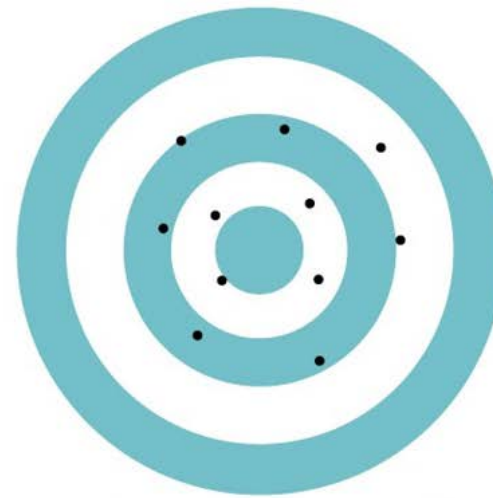
- **En observator er forventningsrett** hvis forventning i dennes utvalgsfordeling er lik sann verdi av parameter. Ellers forventningsskjev.
- **Variabilitet: Spredning i fordeling.** Avhenger av utvalgsstørrelse. Stort utvalg gir mindre variabilitet.

Forventningsskjevhet og variabilitet

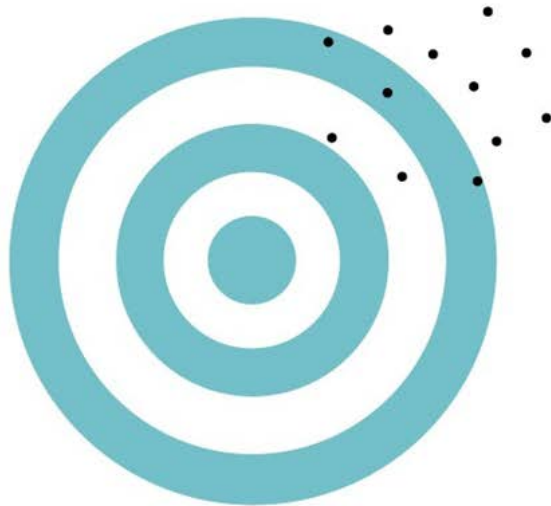
- Vi kan tenke på den sanne parameterverdien for populasjonen som blinken på en skyteskive
 - Observatoren er pilkastet
 - Forventningsskjevhet og variabilitet beskriver hva som skjer når man kaster mange ganger
 - Forventningsskjevhet: Hvor langt fra blinken man systematisk kaster
 - Variabilitet: Hvor spredt man treffer
 - Et godt utvalgs-design og en god pilkaster har begge lav forventningsskjevhet og lav variabilitet



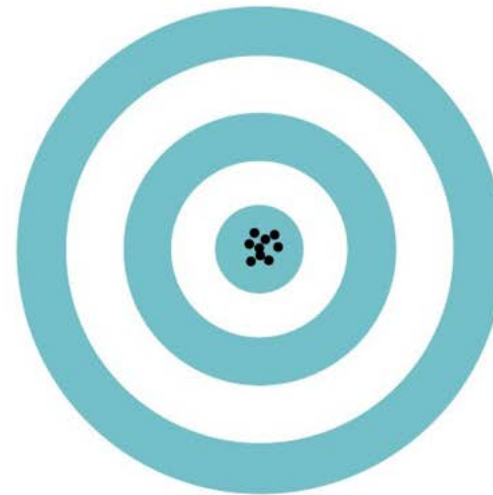
High bias, low variability
(a)



Low bias, high variability
(b)



High bias, high variability
(c)



The ideal: low bias, low variability
(d)

God design

- **Redusere skjevhet:** Tilfeldig utvalg
- **Redusere variabilitet:** Bruk stort utvalg
- **Estimater kommer gjerne med feilmarginer:**
 - Setter grenser for størrelse på feil i estimatet
 - Reflekterer utvalgsvariabiliteten
 - Avhenger av utvalgsstørrelsen (mindre for større utvalg)
- Populasjonsstørrelse betyr ingenting (så lenge populasjonen er minst 100 ganger større enn utvalgsstørrelsen n)
 - Utvalg på 2500 like bra for populasjon med 900.000 individer som for 9.000.000

Hvorfor randomisere

- Fjerne skjevheter, senter i fordeling lik sann parameterverdi
- Kan bruke sannsynlighetsteori for å analysere data
- Form kjent, ofte (nesten) normalfordelt
- Senere vil vi komme tilbake til mer tekniske detaljer for utvalgsfordelinger og hvordan vi kan trekke statistiske konklusjoner basert på dem

DAGENS VIKTIGSTE

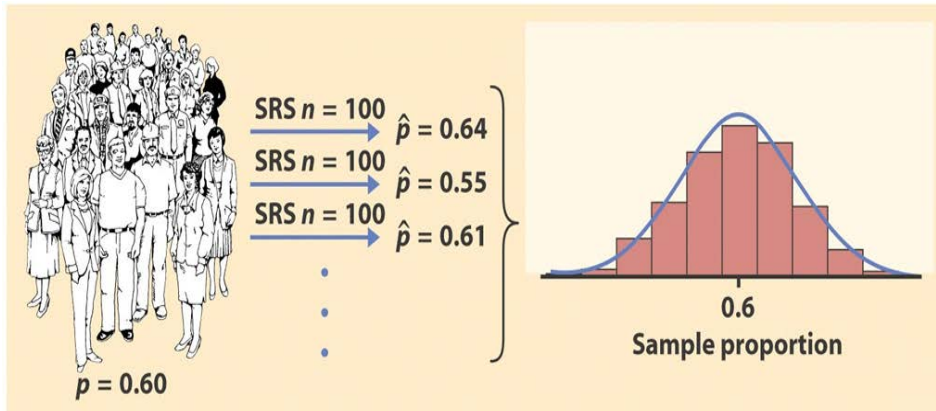


Figure 3-9
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

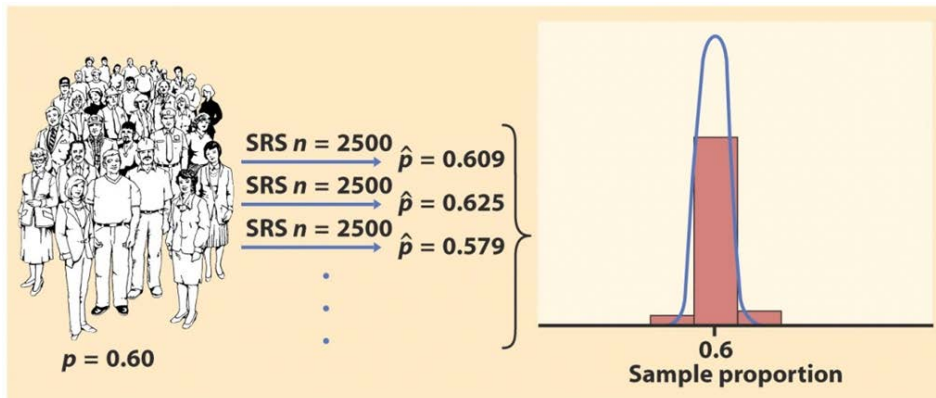


Figure 3-10
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company