

Minitab for skrekkslagne

Espen Andersen

Denne manualen/læreboken ble skrevet sent på 80-tallet, for studenter på BI og andre steder som skulle benytte Minitab, et statistikkprogram som fremdeles er mye brukt.

Jeg begynte forsiktig å oppdatere denne i 1993, men pga. andre arbeidsoppgaver ble jeg aldri helt ferdig. Siden har den ligget i et hjørne av harddisken. En hel del studenter bruker imidlertid fremdeles versjon 2.1 av "skrekkslagne", og spør meg om å få kopiere det ene papir-eksemplaret jeg har, osv. Derfor legger jeg denne litt mer oppdaterte versjonen ut som en PDF-fil, så kan de som har lyst bruke den.

Det understrekes at dette er en kladd – det er en hel del steder teksten ikke er avsluttet, det er rare sidebrekk, innholdsfortegnelsen og stikkordsregisteret stemmer ikke, osv. Jeg kommer ikke til å gjøre noe mer arbeid på den (og ikke spør meg om Minitab detaljer – har glemt det meste....) – men hvis den kan være til nytte – vær så god!

Espen Andersen (espen.andersen@bi.no)
Sandvika, mai 2000

Forord

Denne manualen er ment å være et alternativ for den som ønsker seg en norskspråklig, enkel innføring i bruk av Minitab. Dette er ikke en lærebok i statistikk, men en veiledning og oppslagsbok i hvordan få noe ut av å bruke en relativt utbredt programpakke for statistisk analyse. Opprinnelig er den skrevet for bruk på BIs siviløkonom- og mellomfagsstudium — noe eksempler og emnevalg bærer preg av. Det skulle imidlertid ikke være noe i veien for å benytte den i studieopplegg utenom BI.

Nytt i denne utgaven er utvidet vekt på forskjellige typer regresjonsanalyse (delt in i enkel, multipl og ikke-lineære tilnærminger), faktoranalyse samt endel mer perifert materiale som f.eks. matriseregning og Minitab for Digital VAX-maskiner. Manualen er oppdatert til versjon 8.0 av Minitab for de datamaskintyper den foreløpig finnes for (IBM PC og Apple Macintosh). Imidlertid er det lagt liten vekt på detaljer ved selve brukergrensensnittet (vinduer, menyer, og lignende), siden dette er utmerket forklart i Minitab's egen manualer, og dessuten er kjapt å finne ut av ved litt prøving og feiling.

Endel feil, unøyaktigheter og forglemmelser er forsøkt rettet opp — men som med alle andre lærebøker er det slett ikke umulig at det finnes noen igjen. Skulle leseren finne flere, vil jeg sette pris på å bli gjort oppmerksom på dem, gjerne ved bruk av skjemaet på side 142.

Jeg vil gjerne takke amanuensis Pål Aakre (BI) for gjennomlesning og innsiktsfulle kommentarer (les: retting av de mest graverende feilene). Hans inspirerende forelesninger har "reddet" statistikken for et utall BI-studenter — meg inkludert. Dagfinn Vaksvik (BI) fant på eksemplet med "den enarmede banditten". Professor Richard M. Meyer (Harvard Business School) lærte meg mer komplisert regresjonsanalyse, professor Al Silk (HBS) om krav til validitet ved statistiske undersøkelser. Mette Hoff (BI) har hånd om den praktiske siden av utgivelsen. Knut Jørgensrud, Knut Winsnes og Mari Wolstad (alle BI) har gjort det betydelig enklere for meg å jobbe tvers over Atlanteren. Minitab har støttet meg med oppdateringer og svar på tekniske spørsmål.

Og sist, men ikke minst en takk til min familie — som gjør minst halvparten av jobben ved å tolerere at far og mann er åndsfraværende og stresset i perioder.

Arlington, engang tidlig på 90-tallet.

Innhold

Innledning	6
a. Om Minitab	6
a. Om manualen	7
Grunnleggende om Minitab	10
a. Data slik Minitab ser dem	10
a. Litt om datamaskiners oppbygning og Minitab	11
a. Oppstart og avslutning	12
a. Kommandoer	13
i. Subkommandoer	14
i. For lange kommandoer	15
i. Kommentarer	15
i. Manglende verdi (*)	16
Hjelp!	17
Data inn i — og ut av — Minitab	20
a. Å legge data inn i en ny kolonne	20
a. Å legge data inn i en eksisterende kolonne	21
a. Lagring av data	22
i. SAVE og RETRIEve (worksheets)	23
i. READ og WRITE (datafiler)	24
a. Alfnumeriske variable	26
Manipulering med kolonner	28
a. Å gi kolonner navn	28
a. Å kopiere kolonner	29
i. Mer spesialisert kopiering	30
a. Å slette data eller kolonner	31
a. Å slå sammen flere kolonner til én	31
a. Å dele opp kolonner	32
a. Å forandre data i en kolonne	32
a. Å liste opp innholdet i en kolonne	33
a. Å endre enkelt-tall (LET-kommandoen)	34
i. Funksjoner til LET-kommandoen	35
a. Å sortere en kolonne	39
i. Sortering	39
i. Ranging (sortering uten flytting)	40
Enkle deskriptive kommandoer	41
a. DESCRibe (beskriv)	41
a. Andre beskrivende kommandoer	42
TABLE-kommandoen	44
a. Sub-kommandoer til TABLE	45
Sannsynlighetsfordelinger	51
a. Hvilke fordelinger finnes?	51
a. Å finne sannsynligheter	53
i. Punktsannsynligheter	53
i. Kumulative sannsynligheter	54
a. Å finne kritiske verdier	56
a. Tilfeldige tall og utvalg	56
i. RANDOM-kommandoen	57
i. Et eksempel: De Store Talls Lov	58
i. SAMPLE-kommandoen	59
i. Et simuleringseksempel: enarmet banditt	59
Tester og konfidensintervall	62
a. Z-test og -intervall	62

a. T-test og -intervall.....	64
a. Test av samsvar mellom to utvalg (inferens).....	65
a. Kji-kvadrat-analyser.....	67
a. Variansanalyser.....	69
Grafikk.....	72
a. PLOT (punktdiagrammer).....	73
a. HISTogram.....	74
a. BOXPlot.....	75
a. DOTPLOT (punkthistogrammer).....	76
a. STEM-and-leaf diagram.....	77
Lineær regresjonsanalyse.....	79
a. Enkel lineær regresjon.....	80
a. Multippel regresjonsanalyse.....	82
i. Mer om REGRes-kommandoen.....	82
a. Korrelasjon.....	83
Kontroll av regresjonsanalyse.....	84
a. Statistisk konklusjonsvaliditet.....	85
i. Multikolaritet.....	85
i. Normalitet.....	86
i. Heteroscedastisitet.....	87
a. Intern validitet.....	88
a. Fortolkningsvaliditet.....	89
a. Ekstern validitet.....	89
Avansert regresjonsanalyse.....	90
a. Nyttige triks.....	90
i. Regresjon med indikatorer (dummy-variable).....	90
i. Transformasjon av variable.....	92
i. Gamma-funksjonen.....	92
i. Log-log regresjon.....	93
a. To skumle kommandoer.....	93
i. Å finne de beste prediktorene.....	93
i. Trinnvis multippel regresjonsanalyse.....	94
a. Diskriminantanalyse.....	95
Ikke-parametriske tester.....	98
a. Test av rekkefølge (RUNS).....	98
a. Fortegnstester.....	99
a. Mann-Whitney test.....	100
Utskrift.....	102
a. Utskrift til printer.....	102
a. Utskrift til disk (fil).....	102
a. Continue?.....	103
a. Utskrift fra enkelte kommandoer.....	103
Kommandofiler — å programmere Minitab.....	105
a. Et raskt lite program.	106
a. Hvordan lage kommandofiler.....	106
i. STORE-kommandoen.....	106
i. JOURNAL-kommandoen.....	107
i. Bruk av editorer.....	107
a. Hvordan kjøre kommandofiler.....	108
a. Noen tips.	108
a. Programmeringseksempel: enarmet banditt.....	110
Minitab og andre programmer.....	113
a. Data fra/til regneark.....	113
a. FORTRAN-formaterte data.....	113
a. Minitab og tekstbehandlingsprogrammer.....	115
i. Eksempel: Tekst fra Minitab til Wordperfect 5.x.....	116
Minitab på IBM PC'er.....	118

a. Fullskjerm dataeditor	119
a. Filnavn og -typer (IBM PC)	119
a. Kapasitetsbegrensninger (IBM PC)	120
a. SYSTEM-kommandoen	121
a. Grafikk på IBM PC	121
a. Hvis du ikke har harddisk	123
Minitab på Apple Macintosh	125
a. Spesielle kommandoer (Macintosh)	126
a. Filnavn og -typer (Macintosh)	126
a. Kapasitetsbegrensninger (Macintosh)	127
a. Grafikk på Macintosh	127
Minitab på IBM stormaskin (VM/CMS)	128
a. Lagringsplass og andre standard	128
a. Filnavn og -typer (IBM stormaskin)	129
a. PAPER-kommandoen (IBM stormaskin)	130
a. CMS subset	130
a. Utskrift på skjermen (IBM stormaskin)	130
Minitab på DEC VAX (VMS)	131
a. Lagringsplass og andre standarder	131
a. Filnavn og -typer (DEC VAX)	132
a. VMS-aksess	133
Minitab og matriseregning	134
a. Hvordan løse ligninger med Minitab	134
a. Matrisekommandoer	135
A. Kommentert litteraturliste	137
B. Data brukt i eksempler	140
C. Kommentarskjema	142
D. Stikkordregister	144

Innledning

1.1 Om Minitab

Minitab er et statistikkprogram. Det er billig, enkelt å lære seg og kan kjøres på de fleste datamaskiner. Det er tilstrekkelig kraftig og fleksibelt for undervisning, oppgaver og mindre forskningsprosjekter innenfor anvendt statistikk. På BI benyttes det i grunnleggende statistikkurs på siviløkonom- og revisorstudiet, samt på mellomfagstillegget. Det har vært brukt i en årrekke — fra lenge før dataverktøy ble allemannseie.

Minitab er enkelt å bruke — men til å begynne med ser det ikke slik ut. Dette skyldes at Minitab er *kommandodrevet* og *tekstbasert*. Dette vil si at Minitab kun i begrenset grad gjør bruk av "fancy" grafiske virkemidler som mus, rullegardinmenyer, vinduer o.l. På den annen side ser Minitab nesten lik ut uansett hva slags datamaskin du kjører den på. Dette gjør at hvis du først lærer deg *litegrann* om Minitab, kan du bruke det (nesten) uansett hva slags datamaskin du måtte ha — og uten å måtte lære det om igjen.

Minitab finnes for en rekke datamaskintyper. I denne manualen er behandlet:

IBM-kompatible PC'er (heretter kalt PC). Dette er PC'er slik folk flest kjenner dem — felles for dem er at de benytter operativsystemet DOS (MS-DOS, PC-DOS e.l.). Dette er pr. i dag den mest brukte maskinen blant studenter på BI — og den som er "standardmaskin" i undervisningen. Den første IBM-kompatible PC'en (dvs. IBMs første PC) kom i 1981.

Apple Macintosh (heretter kalt Mac) er en pussig liten boks med sort/hvitt (etterhvert farger på de dyrere modellene) skjerm og en såkalt "mus" som benyttes for å "peke" på de tingene man vil ha utført (på EDB-språk kalles dette "grafisk grensesnitt"). Denne maskinen kom i 1984 — og ble til å begynne med regnet for å være et leketøy. Etterhvert har den vist tenner, og er nå i økende grad den maskinen studenter ønsker seg.

IBM VM/CMS er ikke en datamaskin, men et operativsystem som benyttes på IBM's stormaskiner. Stormaskiner, også kalt mainframes, er store datamaskiner, plassert i egne maskinrom, som mange mennesker benytter samtidig ved å koble seg til ved hjelp av terminaler. Man kan benytte både PC og Macintosh som terminal ved hjelp av et kommunikasjonsprogram (terminalemulator). På BI benyttes en IBM stormaskin av siviløkonom- og revisorstudentene. IBM stormaskiner kan ha andre operativsystemer enn VM/CMS, og er meget vanlige maskiner i større bedrifter.

Da PC Magazine (et av verdens mest kjente datatidsskrifter) i mars 1989 testet 49 statistikkpakker, kom Minitab på en delt førsteplass blant de "grunnleggende" statistikkpakkene (den andre pakken het Statistix). Statistikkpakkene ble delt inn i to klasser (avansert og grunnleggende) og testet grundig. Minitab fikk toppscore pga. sine mange og fleksible kommandoer - samt det faktum at den finnes for mange forskjellige datamaskintyper, og ser nesten lik ut på alle sammen. Nettopp derfor passer programmet utmerket for studenter, som gjerne har tilgang til kraftige datamaskiner av ulike typer, uten nødvendigvis å eie dem selv.

DEC VAX (heretter kalt VAX) er en type flerbrukermaskiner fra Digital Electronic Corporation. Størrelsen varierer, men generelt er de gjerne mindre enn IBM's stormaskiner i kapasitet--til gjengjeld er det vanligere at firma har flere av dem, gjerne koplet sammen i nett (såkalte VAX-clusters) som for brukeren ser ut som en enkelt maskin. Akkurat som med IBM stormaskin finnes det flere operativsystemer for denne maskintypen: VMS (som er Digital egenutviklede operativsystem, og det eneste som er behandlet i denne manualen) og UNIX (som er et operativsystem som i hvert fall i prinsippet skal se likt ut på alle maskintyper som har det). Den er meget vanlig i akademiske miljø og tildels i bedrifter i Norge.

I tillegg finnes Minitab for en rekke andre datamaskintyper (Prime, Hewlett-Packard, Data General osv.), uten at de er behandlet her. Minitab ser nokså lik ut på de forskjellige maskinene — kan man bruke den på én maskin er det lett å bruke den på andre, gjerne ved hjelp av denne manualen. De små forskjeller som finnes har hovedsaklig med hvordan man henter og lagrer data å gjøre, og er her behandlet i egne kapitler.

1.2 Om manualen

Denne manualen er ingen lærebok i statistikk — ei heller i bruk av datamaskiner generelt. Den forutsetter at du er rimelig orientert om hvordan datamaskiner fungerer — hvilket vil si at du kan starte dem, slå dem av, vet hva en fil er, vet hva en diskett og eventuelt en harddisk er — samt er klar over *egentlig* er det ingen heksekunst. (Datamaskiner er i grunnen svært enkle å bruke — men det hjelper på en rekke "eksperters" selvfølelse å la andre tro det motsatte. . .).

Denne manualen kan benyttes på to måter: som *lærebok* (ofte kalt "tutorial" i datasammenheng) eller som *oppslagsbok* (reference). Kapitlene har navn som "data inn i — og ut av — Minitab" eller "enkle beskrivende kommandoer". Hensikten med dette er at du skal kunne finne frem til de funksjoner du ønsker å benytte ved å lese innholdsfortegnelsen. Hvis du vet hva en kommando heter — men lurer på hvordan man skal benytte den — kan du begynne bakerst, med å slå opp i listen over Minitab-kommandoer.

Det er ikke meningen at du skal lese hele manualen—selv om noen sikkert gjør det også —kun de delene du finner interessante. Det anbefales å lese manualen mens du sitter ved siden av en PC eller en terminal. Datamaskiner læres best ved å prøve og særlig ved å feile (omtrent som bilkjøring, men som regel er det ikke like dyrt). Om du leser aldri så mye, vil du vil ikke finne alle Minitabs kommandoer forklart her — kun de som er mest aktuelle for en som skal lære seg statistikk eller analysere mindre datasett. For en komplett oversikt henvises det til litteraturlisten bakerst i manualen. Du kan også finne ut mye ved å bruke HELP-kommandoen.

Kommandoene er forklart slik:

Minitab's røtter

Minitab's historie går tilbake til det som i datamaskinenes historie omtrent tilsvarer den eldre bronsealder. Minitabs utseende (brukergrensesnitt) stammer fra OMNITAB, et interaktivt system som ga en IBM 7090/94 den samme funksjonaliteten som en avansert kalkulator. OMNITAB ble utviklet ca. 1968 ved National Bureau of Standards i USA. OMNITAB hadde kolonner, rader, matriser og konstanter, og refererte til dem på samme måte som Minitab. (Sammett, 1969, pp. 296-299). Minitab ble utviklet for statistikkundervisningen på Penn State University. Ettersom etter-spørselen utviklet seg, ble det dannet et eget firma, som siden har solgt Minitab, primært til skoler og universiteter, men også til større firma som sysler med tidsserie-analyser for produksjonsplanlegging m.m.

KOMMando [x] kolonne(r)

for eksempel:

SET kolonne ['filnavn']

Dette betyr at kommandoen SET *skal* ha med en kolonne — og den *kan* ha med et filnavn. Ord som står innenfor [hakeparenteser] er ting du *kan*, men ikke *må* ha med. Alle Minitabs kommandoer kan forkortes ned til 4 bokstaver — derfor har jeg skrevet dem med kun disse fire bokstavene som versaler (store bokstaver). Når ting står skrevet i *kursiv*, betyr dette at du ikke skal skrive ordrett det som står der — men i stedet bytte ut med det som er aktuelt (i eksemplet ovenfor skal man ikke skrive "filnavn" — men i stedet navnet på filen).

I teksten finnes mange utlisteringer av Minitab-sesjoner. Her er alle kommandoer o.a. som skrives inn fra tastaturet skrevet med **uthevet skrift**, mens det som maskinen skriver er gjengitt med *vanlig skrift*.

En rekke emner er som sagt ikke behandlet, eller ikke behandlet grundig. Dette er gjort dels ut fra en vurdering av hva som er sentralt i forhold til de forventede lesere, dels ut fra personlige preferanser, dels ut fra tidshensyn. Av emner som ikke er behandlet her finner vi

- tidsserier: dette er analyseteknikker som går ut på å studere utviklingen av en prosess (som regel en produksjonsprosess) over tid
- statistisk kvalitetssikrings: dette er teknikker som benyttes for å analysere og sikre kvalitetsnivåer (quality analysis/quality control). Dette er et relativt nytt sett kommandoer til Minitab, aktuelt innenfor materialadministrasjon.
- EDA (exploratory data analysis): dette er endel kommandoer som benyttes til å få oversikt over datasett før analyse. Noen av disse kommandoene (BOXPlot, STEM-and leaf) er med under andre punkter i manualen.

Av ting som kun er nevnt kort finner vi

- FORTRAN-formaterte data
- alfanumeriske data
- ikke-parametrisk statistikk (endel tester er ikke tatt med)
- variansanalyse (kun "oneway" ANOVA er behandlet, og kun med et eksempel)
- høyoppløselig grafikk
- diskriminantanalyse
- PCA (principal component analysis, som er en form for faktoranalyse)

Minitab mangler endel statistiske tester som begynner å bli mer vanlig i bruk, særlig etterhvert som datamaskiner stadig blir billigere og kraftigere. Blant disse metodene vil jeg nevne

- binær regresjon (LOGIT/PROBIT analyse), dvs. regresjonsanalyse hvor den avhengige variabelen (Y) er enten 0 eller 1. Binær regresjonsanalyse finner sannsynligheter for at Y skal ha en av disse to verdiene. Det finnes en makro (et "program" for Minitab) som kan utføre en enkel variant av LOGIT/PROBIT.
- regresjonsanalyse hvor den avhengige variabelen (Y) er begrenset av ytre forhold (TOBIT-analyse). Et eksempel her kan være salg av kinobilletter — oppad begrenset av hvor mange

seter det er i kinoen. Etterspørselen etter billetter kan være mye høyere enn antall seter, men så lenge man kun har data for antall billetter solgt, vil vanlig regresjonsanalyse gi et galt bilde av sammenhengen mellom pris og etterspørsel.

Jeg har forsøkt å benytte norske betegnelser der jeg mener det er naturlig. Som regel er de engelske uttrykkene satt i parentes ved siden av — dels fordi du vil støte på dem i annen dokumentasjon og i Minitabs hjelpetekster, dels fordi de norske uttrykkene ikke er direkte oversettelser og derfor kan skape forvirring. Jeg har imidlertid ikke gått til slike eksesser som å skrive "PD" (Personlig Datamaskin) i stedet for PC, platelager i stedet for harddisk osv. Både innenfor statistikk og informasjonsteknologi er arbeidsspråket ofte engelsk — og frenetisk fornorsking vil da ofte bare virke sært.

.og dermed skulle vi kunne begynne.

Grunnleggende om Minitab

I dette kapitlet skal vi få en oversikt over Minitab — hvordan Minitab behandler og lagrer data, hvordan du gir kommandoer for å behandle dataene, og hvordan du starter og avslutter Minitab.

2.1 Data slik Minitab ser dem

Vanligvis benytter man Minitab til å *analysere* data man har samlet sammen på forhånd. Disse dataene kan være nær sagt hva som helst — et enkelt dataelement kalles gjerne en *observasjon*. En observasjon kan f.eks. være at du har målt høyden på en person og funnet at han eller hun er 176 cm høy. Hvis vi måler høyden på 3 gutter og 3 jenter, kan vi få dette resultatet:

gutt 189 cm, jente 165 cm, gutt 172 cm, jente 157 cm, gutt 181 cm, jente 174 cm

Hvis vi så tenker oss at vi i stedet for "gutt" og "jente" skriver henholdsvis 1 og 2, og at vi setter opp resultatet i to *kolonner*, får vi følgende:

1	189
2	165
1	172
2	157
1	181
2	174

Dette er måten Minitab arbeider med data; lagret i *kolonner* (columns) som representerer *variable* (her: kjønn og høyde). Hver *rad* (row) representerer en person — dvs. en *observasjon* eller, hvis det er snakk om en data fra en spørreundersøkelse; en *respondent*.

Kolonnene i Minitab er nummerert: C1, C2 osv. Du velger selv nummeret. Antallet kolonner det er lov å bruke varierer fra maskintype til maskintype: på PC og Mac er det lov å ha inntil 100 kolonner samtidig, på stormaskin 1000.

Dataene må i hovedsak lagres som tall (med visse, noe kronglete unntak). Dette betyr at du må **kode** data som ikke fra før av består av tall (f.eks. gutt/jente, ja/nei, galt/riktig, fylke, bilmerke el.l.).

Tall i Minitab kan være heltall (engelsk: integers) eller desimaltall. I stedet for komma brukes **desimalpunktum**, på engelsk/amerikansk manér. Dette vil si at "åtte komma nitti" skrives slik: 8.90.

Du kan også lagre data i Minitab i form av konstanter og matriser.

Konstanter (constants) betegnes med K1, K2 osv., og inneholder et enkelt tall. Tre konstanter er forhåndsdefinert (dvs. de inneholder en verdi med mindre man putter noe annet i dem):

K98 inneholder tegnet "*" (stjerne) som i Minitab betyr "manglende verdi" (missing value).

K99 inneholder tallet e (grunntallet i det naturlige logaritmesystem, omtrent 2.718281828459045. . .).

K100 inneholder tallet π (pi, dvs. forholdet mellom en sirkels omkrets og dens diameter, omtrent 3.141592653589793. . .).

Matriser betegnes med M1, M2 osv. Også her varierer det maksimale antallet med hvilken datamaskintype du benytter.

2.2 Litt om datamaskiners oppbygning og Minitab

(Den "proffe" databruker kan hoppe over dette punktet. . .).

Alle datamaskiner har en sentralenhet (CPU, Central Processing Unit). På en PC er det den delen vi står igjen med når vi har fjernet skjermen, tastaturet og evt. skriveren. CPUen består igjen av tre enheter: prosessoren, det indre (interne) og det ytre (eksterne) lageret.

Dette kan synes noe komplisert — men la sammenligne sentralenheten på en datamaskin med en person som sitter på et kontor, med et skrivebord og et arkivskap.

Prosessoren er regneenheten: den leser data fra indre lager, behandler/beregner data, leser det ut igjen til det indre lageret. I vår sammenligning med en kontorarbeidsplass tilsvarer altså prosessoren den personen som sitter og arbeider.

Det indre lageret kalles ofte internhukommelsen. En annen betegnelse er RAM (Random Access Memory). Internhukommelsen fungerer som et "nærlager" for prosessoren — der legges data som det er viktig å få tak i kjapt. På den annen side er dette lageret av begrenset størrelse, og det som ligger lagret der forsvinner når datamaskinen slås av. I "kontormaskinen" vil vi kunne sammenligne internhukommelsen med skrivebordet — det er av begrenset størrelse — men det som ligger der er de tingene man arbeider mest med og som man ønsker å kunne få tak i fort.

Det ytre lageret kan være av forskjellig slag — for brukere av PC er de to mest aktuelle diskett og/eller harddisk. Dette lageret er mer robust enn internhukommelsen (det forsvinner ikke når man slår av strømmen) og det er større (for diskettens vedkommende er det jo ubegrenset — man kan ha så mange disketter man vil. . .). Imidlertid går det mye tregere for prosessoren å få tak i noe fra det ytre lageret. (Harddisk, som kan sies å være en stor diskett fast montert inne i PC'en, er imidlertid adskillig raskere enn disketter — men ikke på langt nær så raskt som internhukommelsen). Hvis vi igjen sammenligner med et kontor, er det ytre lageret det samme som et arkivskap — det kan være meget stort, man kan lagre masse informasjon som gjenfinnes etter systematiske regler.

Hvilken betydning har så dette når du skal benytte Minitab?

Når du legger data inn i kolonner, gjør beregninger på dem osv., lagres disse kolonnene i datamaski-

nens internhukommelse. Det området av internhukommelsen som Minitab benytter til å lagre data som brukes i, kalles *arbeidsområdet*. Hvis du går ut av Minitab, blir de slettet. Derfor er det viktig at du lagrer disse dataene til et mer permanent lager før du avslutter Minitab og/eller skrur av maskinen. Dette gjør du med kommandoene SAVE eller WRITE (hvilken av dem du bruker er avhengig av hva du skal bruke dataene til etterpå). Disse kommandoene flytter dataene fra internhukommelsen over til det ytre lageret (disketten eller harddisken) og legger det der som en *fil*. Filen har et navn som er unikt — dvs. at du ikke kan ha to filer på det samme lagringsmediet med det samme navnet. Du kan kalle filen hva du vil innenfor visse regler (disse reglene varierer ut fra hva slags datamaskin du har).

2.3 Oppstart og avslutning

Hvordan Minitab skal startes avhenger av maskintype. Som regel skriver man "Minitab" og trykker retur-tasten (på Macintosh "dobbelklikkes" det på Minitab-symbolet), etter først å ha startet eller logget seg på maskinen. Når Minitab først er startet, er den stort sett lik på alle datamaskiner (se forøvrig kapitlene om Minitab på ulike maskintyper; kapittel ?, ?, ? og ?).

Når du har startet Minitab, vil du etter en stund få teksten

```
MTB >
```

på skjermen. Dette er Minitabs "prompt" (klartegn), og det betyr "Minitab er klar og venter på en kommando".

For å avslutte Minitab, skriver du

```
STOP
```

og trykker Enter/Return-tasten (på Macintosh velger du Quit fra File-menyen).

NBNBNBNB!

I motsetning til de fleste andre dataprogrammer gir ikke Minitab deg noen advarsel hvis du ikke har lagret dine data til disk!

Bruk derfor *alltid* SAVE- eller WRITE-kommandoen (s.d.) hvis du skal spare på de dataene du har arbeidet med!!!!

Kom bare ikke her og si du ikke var advart. . . .

2.4 Kommandoer

Minitab er et *interaktivt* program. Dette betyr at man arbeider med Minitab ved å gi *kommandoer*, som Minitab utfører, for så å gi tilbakemelding om resultatet straks kommandoen er utført¹.

¹Det motsatte av interaktiv databehandling er *satsvis* (batchorientert). Dette betyr at man skriver sine kommandoer i et program, som lagres som en fil, og så startes. En del "proffere" statistikkpakker som SAS (Statistical Analysis System) og SPSS (Statistical Package for the Social Sciences) er hovedsaklig batchorientert.

En kommando gir du ved å skrive den etter Minitabs prompt (klartegn). Du trenger ikke skrive mer enn de fire første bokstavene av en kommando (du kan godt skrive den helt ut allikevel — greit f.eks. hvis du skal benytte kommandoene til en rapport, fagoppgave el.l.). La oss ta et eksempel:

Kommandoen **INFORMATION** gir en oppsummering av hvilke kolonner, konstanter og matriser som er i bruk. Denne kommandoen gis slik (du skriver kun det som er **uthevet**):

```
MTB > I N F O                                (kommandoen gis)
      * ALL COLUMNS EMPTY AND UNNAMED (Minitab svarer. . .)
```

```
CONSTANTS USED: NONE
```

```
MTB >                                           (. . .og er klar til ny kommando)
```

Vi gjentar: Som du ser, utfører Minitab kommandoen med en gang (og svarer, i dette tilfelle, at det ikke ligger noen data i arbeidsområdet). Når kommandoen er utført, kommer *MTB >* frem igjen — dvs. Minitab er klar for neste kommando.

Man kan også skrive kommandoene slik:

```
I N F O R M A T I O N
i n f o r M a t I o n
```

eller for den saks skyld

```
I N F O h v a s o m h e l s t
```

Minitab leser som sagt kun de 4 første bokstavene i hver kommando — og lager ingen problemer med resten *såfremt linjen ikke inneholder noe som kan tolkes som kolonner, tall el.l.*

2.4.1 Subkommandoer

Mange av Minitabs kommandoer har *subkommandoer*. Dette vil si "tilleggskommandoer" eller spesifikasjoner av hvordan hovedkommandoen skal utføres.

En subkommando gis på følgende måte:

- etter hovedkommandoen setter du et semikolon (;)
- Minitab svarer med klartegnet "*SUBC>*"
- du skriver subkommandoen
- hvis det ikke er flere subkommandoer avslutter du subkommandoen med et punktum.
- hvis det er flere subkommandoer avslutter du subkommandoen med et semikolon, for deretter å fortsette på neste linje

Forvirret? La oss ta et lite eksempel igjen (her brukes **COPY**-kommandoen, den kan du lese mer detaljert om på side 29):

```
MTB > COPY C4 C14;                            (Kopier C1 til C14, men. . . .)
      SUBC> USE 20: 60.                          (. . .kun f.o.m. rad 20 t.o.m. rad 60)
```

MTB > (Kopieringen er utført)

I stedet for å avslutte siste subkommando med et punktum, kan du sette et semikolon, og på neste linje skrive END.

```
MTB > COPY C4 C14; (Kopier C1 til C14, men. . . .)
SUBC> USE 20: 60; (. . .kun f.o.m. rad 20 t.o.m. rad 60)
SUBC> END (Alternativt kan du bruke END i stedet for punktum
etter siste subkommando)
MTB > (Kopieringen er utført)
```

Hvis du verken setter semikolon eller punktum, antar Minitab at det skal være semikolon, skriver ut en feilmelding, og lar deg fortsette.

```
MTB > COPY C4 C14; (Kopier C1 til C14, men. . . .)
SUBC> USE 20: 60 (. . .glemte semikolon her!)
* Subcommand does not end in . or ; (; assumed)
(Feilmelding fra Minitab)
SUBC> END (Alternativt kan du bruke END i stedet for punktum
etter siste subkommando)
MTB > (Kopieringen er utført)
```

Hvis du skriver en feil og derfor ønsker å avbryte hele kommandoen, kan du gi subkommandoen ABORT. Minitab avbryter da **hele** kommandoen, og svarer med "MTB >"-klartegnet.

```
MTB > COPY C4 C13; (Au da, skulle vært C14. . . .)
SUBC> ABORT (. . .avbryter hele kommandoen)

MTB > (Klar for ny kommando)
```

Det er ikke nødvendig å skrive punktum etter verken END eller ABORT.

2.4.2 For lange kommandoer

Hvis en kommando (eller en linje med data) går over flere linjer, kan du fortsette den ved å skrive "&" ("grisehale") og fortsette på neste linje. Minitab vil forandre sin "prompt" (klartegn) fra "*MTB* >" til "*CONT*>" (for continue) og la deg fortsette å skrive.

Dette skjer ikke så ofte med kommandoer fra tastaturet — men det er svært vanlig når du skal lese inn data fra datafiler.

2.4.3 Kommentarer

Hensikten med kommentarer er å gjøre utskriften av det du har gjort mer forståelig. Du kan skrive kommentarer på to måter:

- som egen kommando (NOTE) slik:

```
MTB > NOTE Dette er en kommentar
```

- som kommentar *etter* den vanlige kommandoen. Du må da sette #-tegnet (kan variere mellom ulike typer datamaskiner (se kapittel ?, ?, ? og ?)) foran kommentaren, slik:

```
MTB > COPY C1 C2 # Kommentar settes her
```

- som vanlig tekst inne i kommandoen. I og med at Minitab kun leser de fire første bokstavene i hver kommando, samt tall og tegn som kolonner, semikolon o.l., kan dette ofte "klargjøre" hva du gjør:

```
MTB > COPY data fra kolonne C1 til kolonne C2
```

Vær imidlertid litt forsiktig: for det første kan en slik kommentar få en kommando til å se ut som den gjør noe annet enn det den faktisk gjør — for det andre kan du komme til å putte inn noe som er ment som en kommentar — men som faktisk forandrer kommandoens betydning.

2.4.4 Manglende verdi (*)

Innenfor statistikken (og i mange andre sammenhenger også) får du ofte behov for verdien "manglende" eller "ikke observert". Dette kan f.eks. skje hvis vi skal analysere svarene på et spørreskjema — og respondentene (de som skal svare på skjemaet) ikke har svart på alle spørsmål. Det holder ikke å bruke et tall (f.eks. 0) til å betegne dette — rett og slett fordi null også er en verdi — som får betydning f.eks. ved beregningen av gjennomsnitt eller konfidensintervall.

I Minitab benyttes symbolet

*

(asterisk, stjerne) for "manglende verdi" (missing value). Dette behandles som om det skulle være et vanlig tall i alle kommandoer som leser data inn eller ut av Minitab. De fleste kommandoer som gjør statistiske beregninger (f.eks. regresjonsanalyse) fjerner "missing value" fra sine beregninger.

Hvis du har alfanumeriske variable (dvs. både bokstaver og tall) i en kolonne, er koden for manglende observasjon et mellomrom (space).

Hjelp!

Som de fleste andre dataprogram har Minitab en hjelpefunksjon. HELP i Minitab er meget god. så god at det nesten ikke er nødvendig å ha noen manual til Minitab — forutsatt at du er rimelig engelskkyndig og har "forskernese".

Kommandoen for å få hjelp er (naturligvis)

HELP

Hvis du lurer på hvordan du skal bruke HELP-kommandoen, skriver du

HELP HELP

Hvis du lurer på hva Minitab kan gjøre, kan du gi kommandoen

HELP OVERVIEW

Du får da ut denne teksten:

*General help is available in the categories below. Type
HELP OVERVIEW followed by the appropriate number. For
example:*

HELP OVERVIEW 1

<i>1 Introduction to Minitab</i>	<i>6 Missing Value Code</i>
<i>2 Syntax of Commands</i>	<i>7 Files and Devices</i>
<i>3 Subcommands</i>	<i>8 Computer and Local Details</i>
<i>4 Stored Constants</i>	<i>9 Documentation</i>
<i>5 Matrices</i>	<i>10 Versions of Minitab</i>
	<i>11 Alpha Data</i>

Hvis du f.eks. skriver

HELP OVERVIEW 8

får du opp en tekst som forklarer detaljer om Minitab på den datamaskintypen du bruker (computer and local details, dvs. spesifikasjoner av hvor store datasett du kan benytte m.m.).

HELP kan også gi deg en liste over alle tilgjengelige kommandoer. Dette gjøres med

HELP COMMANDs

Du får da frem denne teksten:

1 General Information	11 Multi vari ate Analysis
2 Entering Data	12 Nonparametri cs
3 Outputting Data	13 Tables
4 Editing and Manipulating Data	14 Time Series
5 Arithmetic	15 Exploratory Data Analysis
6 Column and Row Operations	16 Distributions & Random Data
7 Plots and Histograms	17 Sorting
8 Basic Statistics	18 Matrices
9 Regression	19 Miscellaneous ²
10 Analysis of Variance	20 Stored Commands and Loops
	21 How Commands are Explained in HELP

For å få en liste over alle kommandoer som vedrører regresjonsanalyse, skriver du altså

HELP COMMANDs 9

Hvis du trenger hjelp til en bestemt kommando, skriver du

HELP kommando

f.eks.

HELP SORT

Hvis hjelpeteksten er lang, kan den være delt opp i sekvenser. Dette vil si at Minitab skriver ut første "porsjon" av hjelpeteksten (som regel en skjerm full), for så å spørre

Conti nue?

Du kan da svare N for NO hvis du har fått svar på det du spurte om, eller bare trykke Retur-tasten hvis du ønsker mer informasjon. Hvis du ønsker å få *hele* hjelpeteksten skrevet ut (selv om den går over flere skjermbilder) kan du gi denne kommandoen:

HELP kommando +

Dette kan være greit hvis du benytter det sammen med OUTFile-kommandoen (se pkt. ?, side 102) — da kan du skrive ut hjelpeteksten på papir og studere den i ro og mak etterpå.

Minitab lister opp en kommandos subkommandoer i hjelpeteksten — men forklarer sjelden subkommandoene. Hvis du trenger hjelp med en subkommando, kan du gjøre det slik:

HELP kommando subkommando

f.eks.

²Miscellaneous: Engelsk uttrykk for "ting som ikke kan settes i egen kategori"

HELP TABLE ROWPercent

Data inn i — og ut av — Minitab

I dette kapittelet skal vi se hvordan vi kan legge data inn i Minitab (noe som i hovedsak gjøres med READ-kommandoen), hvordan vi kan lagre data i en fil — og hvordan vi kan hente dem (eller andre data) inn igjen.

4.1 Å legge data inn i en ny kolonne

For å lese data inn i en *enkelt* kolonne bruker vi SET-kommandoen, ved å skrive

```
SET kolonne(r)
```

Minitab forandrer prompt (klartegn) fra "*MTB >*" til "*DATA>*". Dette betyr at Minitab venter på data. Vi skriver dataene, atskilt med mellomrom (space) eller komma. Når vi ikke har flere data, skriver vi END. Her er et eksempel:

```
MTB > SET C1          ("Åpner" kolonne C1)
DATA> 34 36 34.5 35    (Data legges inn med mellomrom. . . .
DATA> 28, 36, 36, 33  . . . eller med komma imellom)
DATA> END             (Det var det, ikke mer data)
MTB > PRINT C1        (Skriv ut kolonnen på skjermen)
```

```
C1
      34      36      34.5      35      28      36      36      33
```

```
MTB >
```

Denne kolonnen ligger nå i *arbeidsområdet* (dvs. i en del av maskinens internhukommelse).

Hvis mange av tallene er like og følger etter hverandre, har SET-kommandoen visse muligheter for å la deg putte in mange tall samtidig. For eksempel:

```
MTB > SET C2
DATA> 6(34)
DATA> END
MTB > PRINT C2
```

```
C2
      34      34      34      34      34      34
```

```
MTB >
```

Dette kan gjøres mer avansert — her er noen eksempler:

6: 10	gir tallene	6, 7, 8, 9, 10
4: -1	gir tallene	4, 3, 2, 1, 0, -1
0: 10/3	gir tallene	0, 3, 6, 9
1: 3/. 5	gir tallene	1, 1.5, 2, 2.5, 3
3(1, 2, 4)	gir tallene	1, 2, 4, 1, 2, 4, 1, 2, 4
(1, 2, 4)3	gir tallene	1, 1, 1, 2, 2, 2, 4, 4, 4
3(1, 4)2	gir tallene	1, 1, 4, 4, 1, 1, 4, 4, 1, 1, 4, 4

NB! Det er viktig at det ikke er mellomrom mellom tallene utenfor parentesene og parentesene.

Eksperimenter litt med dette — det kan spare deg for mye "puncharbeid". . . .

READ-kommandoen omtrent som som SET — med unntak av at den lar deg legge inn i *flere kolonner samtidig*. Syntaks er

READ *kolonner*

Som skilletegn mellom kolonnene benytter du enten mellomrom (ordskiller) eller komma. F.eks vil

```
MTB > READ C1 - C4
DATA> 11 12 13 14
DATA> END
```

gi fire kolonner med tallene 11, 12, 13 og 14 i. For at dette skal fungere, må det være like mange dataenheter som det er kolonner.

4.2 Å legge data inn i en eksisterende kolonne

Hvis du skal legge data inn i en eller flere kolonne(r) som allerede inneholder data, kan du ikke benytte READ- eller SET-kommandoen; den sletter nemlig de data som finnes i kolonnen(e) fra før. I stedet kan du bruke INSErt-kommandoen. Denne kommandoen fungerer på samme måte som READ-kommandoen, med unntak av at den legger data til en kolonne — og den må i motsetning til SET-kommandoen ha beskjed om mellom hvilke rader i kolonnen dataene skal legges inn.

Kommandoen gis altså slik:

INSErt *rad1 rad2 kolonne*

eller, for å ta et eksempel:

```
MTB > INSErt 26 27 C3 C4
DATA> 56 170
DATA> 75 176
DATA> 80 189
DATA> END
      3 ROWS READ
```

som putter inn tre rader med data mellom rad 26 og 27 i kolonnene 3 og 4.

Du kan også benytte en fil for å lese inn dataene i stedet for tastaturet, på samme måte som med READ-kommandoen. F.eks. slik:

```
MTB > INSErt 'FIL1' 26 27 C3 C4
```

Som vanlig må filnavnet ligge innenfor apostrofer.

Hvis du ikke tar med linjenumre, legger INSErt-kommandoen dataene til slutt i kolonnen(e). For å legge in data helt øverst i en kolonne, benytter du (naturligvis) linjenumrene 0 og 1.

4.3 Lagring av data

Det er tre måter lagre data på i Minitab: som *worksheets*, *portable worksheets* eller som *datafiler*. Hver av måtene har sine fordeler og ulemper.

Worksheets er filer som er lagret i et spesielt Minitab-format. Dette formatet er forskjellig for ulike maskiner — dvs. at du ikke kan lagre et worksheet på en IBM PC og flytte det direkte over til en Macintosh eller stormaskin. Du kan heller ikke flytte dataene over i andre programpakker, selv om de er på den samme typen maskin. Imidlertid har worksheets den fordelen at alle kolonner, konstanter og matriser blir lagret akkurat slik de er (inklusive navn på kolonner og matriser). Dermed kan du lett beholde oversikten over dataene når du henter dem inn igjen neste gang du bruker Minitab. Worksheets lagres med SAVE-kommandoen, og hentes inn med RETRIEve-kommandoen. Det er ikke mulig å skrive ut worksheets på en skriver — da må de hentes inn i Minitab og lagres som datafiler først.

Portable worksheets er filer lagret i et format som er spesielt for Minitab, men ikke for maskintyper. Det betyr at hvis du lagrer et worksheet i dette formatet, kan du overføre filen til en annen type datamaskin (f.eks. fra en PC til en Macintosh) og hente den inn i Minitab der. Du beholder da alle detaljer som konstanter, kolonnenavn osv. Ulempen med formatet er at det tar mye større plass enn Minitabs vanlige worksheets, og er langsommere å hente inn og skrive ut.

Datafiler er filer i standard tekstformat. Dataene lagres som kolonner atskilt av mellomrom (ordskiller, space) eller komma. En datafil kan flyttes over til andre datamaskiner eller programpakker — noe som kan være greit hvis du skal inkludere data i rapporter, eller flytte dem til/fra et regneark eller lignende. Imidlertid kan du ikke lagre betegnelser som kolonnenavn o.l. i en datafil — og du kan heller ikke lagre kolonner, matriser og/eller konstanter sammen i en fil. Datafiler lagres med WRITE-kommandoen, og hentes inn med READ-kommandoen. En datafil kan du som regel skrive ut på en skriver uten problemer (litt avhengig av hva slags datamaskin og skriver du har).

Minitab for IBM-kompatible PC'er har også muligheten til å lagre i Lotus 123-format (se kapittel ?, side 118).

4.3.1 SAVE og RETRIEve (worksheets)

Kommandoen for å lagre de dataene du holder på med som et worksheet er

```
SAVE ['filnavn']
```

Filnavnet kan du velge selv — med visse begrensninger avhengig av hva slags datamaskin du bruker. Hvis du ikke oppgir noe filnavn, bruker Minitab et standard-filnavn; hva dette er avhenger også av hva slags datamaskin du bruker.

Ved SAVE-kommandoen lagres **alle** data du har i de ulike kolonner, konstanter og matriser — inklusive navn på kolonner m.v.

For å hente inn et worksheet gir du kommandoen

```
RETRieve ['filnavn']
```

Du kan bare hente inn worksheets med RETRieve-kommandoen (dvs. kun data som har blitt lagret med SAVE-kommandoen). Hvis du ikke spesifiserer noe filnavn, hentes et worksheet med samme filnavn som under SAVE-kommandoen (se ovenfor).

Hvis du ønsker å lagre eller hente et worksheet i et format som kan hentes inn i Minitab for en annen maskintype enn den du har, kan du gjøre dette med subkommandoen PORTable. Dette kan f.eks. gjøres slik:

```
MTB > SAVE 'mi nedata' ;  
SUBC> PORTable.
```

RETRieve-kommandoen fungerer på samme måte. PORTable-formatet finnes kun fra og med Minitab versjon 7.1. Et problem med PORTable-formatet er at det tar forholdsvis mye plass.

4.3.2 READ og WRITE (datafiler)

READ-kommandoen brukes mest for å hente inn data fra filer inn i kolonner. WRITe lar deg skrive data fra kolonner til en fil. Filene som benyttes, ser ut som kolonner med tall. Hver kolonne er atskilt med et mellomrom eller et komma.

Fordelen med å bruke READ- og WRITe-kommandoen er, i tillegg til at filene til en viss grad kan benyttes i andre programpakker enn Minitab, at du kan hente inn eller skrive ut **noen** kolonner — men du trenger ikke lese inn eller skrive ut **alle**. Hvis du benytter SAVE eller RETRieve-kommandoen lagres eller leses **alle** kolonner til eller fra disk, og du mister de du evt. måtte ha i bruk før du ga RETRIEVE-kommandoen.

La oss se litt detaljert på READ- og WRITe-kommandoen, ved hjelp av et eksempel:

```
MTB > READ C1 C2  
DATA> 2. 34 355 (legg merke til desimalpunktum)  
DATA> 5, 123 (komma kan også skille kolonner)  
DATA> . 26 (Auda — kun en kolonne. . . .)  
* ERROR * INCOMPLETE ROW - REENTER (feilmelding. . . .)  
DATA> . 26 56 (trenger ikke null foran desimalpunktum)
```

```

DATA> 3. 7*          ("*" betyr "manglende observasjon")
DATA> END           (END-kommandoen avslutter innlesningen)
  4 ROWS READ       (Minitab svarer antall rader)
MTB > _            (. . .og det var det)

```

Her er READ-kommandoen benyttet til å lese inn dataene fra tastaturet. Som du ser, må READ ha beskjed om hvor mange kolonner som skal "fylles" — og den må ha kolonnene radvis, med like mange tall som det er kolonner.

Hvis du skal lese inn dataene fra en fil, må du først ha laget filen. Dette kan gjøres på mange måter — avhengig av hva slags datamaskin du har. (På en PC kan du f.eks. benytte WordPerfect eller et annet tekstbehandlingsprogram og lagre filen i DOS-format). En annen mulighet er å benytte et editeringsprogram, eller bruke et regneark (f.eks. Lotus 123 eller Excel) og skrive ut filen til disk.

Uansett hva slags program du benytter, skal filen se slik ut for at resultatet skal bli som i eksemplet ovenfor:

```

2. 34 355
5 123
. 26 56
3. 7 *

```

Du kan godt ha **flere** mellomrom mellom tallene (men ikke flere kommaer, hvis du bruker kommaer som skilletegn). La oss si at du har lagret denne filen med navnet "FIL1". Kommandoen for å lese den inn i kolonne 1 og 2 blir da

```

MTB > READ 'FIL1' C1 C2    (filnavnet i apostrofer)
  4 ROWS READ

```

ROW	C1	C2	
1	2. 34	355	(Minitab skriver ut resultatet)
2	5. 00	123	
3	0. 26	56	
4	3. 70	*	

```

MTB > _

```

Hvis det er mange rader, skriver Minitab bare ut de 5 første, samt opplyser om hvor mange rader som er lest inn. Hvis det er feil i en rad (som i eksemplet uten fil ovenfor) avbryter Minitab innlesningen med en feilmelding — uten å lese inn noen data i kolonnene.

SET-kommandoen har to subkommandoer: FORMat og NOBS. FORMat går på spesifisering av hvordan dataene ser ut — se eget punkt om FORTRAN-formaterte data, side 113. NOBS betyr "number of observations", og spesifiserer hvor mange observasjoner (rader) som skal leses inn med SET, READ eller INSErt-kommandoen. Den kan være grei å bruke hvis du ikke ønsker å lese inn alle dataene fra en fil — men kun f.eks. de 100 første. Eksempel:

```

MTB > READ 'DATAFIL' C1-C3;
SUBC> NOBS 100.

```

Her blir kun de 100 første radene lest inn. Hvis det er færre enn 100 observasjoner (rader) i filen, blir NOBS-kommandoen ignorert. NOBS brukes også ved makro-programmering (se kapittel ?, side 105).

WRITE-kommandoen benyttes som en READ-kommando — men "omvendt", i den forstand at kommandoen skriver data ut til en fil i stedet for å lese dataene inn i arbeidsområdet. Hvis du gir kommandoen

```
WRITE 'FIL1' C1 C2
```

blir innholdet i kolonnene C1 og C2 skrevet ut til filen FIL1 (filnavn og -type varierer noe mellom ulike type maskiner). Dataene ligger fremdeles i arbeidsområdet (maskinens internhukommelse). Hvis filen finnes fra før, vil Minitab spørre

```
Overwrite existing file (Y/N):
```

Du svarer da Y(es) hvis du vil slette den gamle filen og erstatte den med de nye kolonnene, eller N(o) hvis du ønsker å avbryte for eventuelt å finne et nytt filnavn.

4.4 Alfnumeriske variable

Uttrykket *alfanumerisk* betyr "tall eller bokstav" — i Minitab er en alfanumerisk variabel en variabel som består av bokstaver (eventuelt også tall) i stedet for bare tall. Du kan lagre opplysninger som alfanumeriske variable i Minitab — men med relativt sterke begrensninger på hva du kan gjøre med dem. I praksis blir alfanumeriske variable sjelden brukt — selv har jeg brukt Minitab siden 1983 og aldri brukt dem (men ofte ønsket at de var litt lettere å bruke).

Alfanumeriske variable kan være inntil 80 tegn lange. De legges enklest inn med dataeditoren (som foreløpig kun finnes i PC-versjonen av Minitab, se kap. ?, side 118), eller ved "formatert" input (side 113). Du kan sortere kolonner etter alfanumeriske variable — men du kan ikke tabulere eller benytte andre beskrivende kommandoer i noen særlig utstrekning. Hvis du skal gjøre dette, må du kode om dataene slik at du får en tallvariabel i stedet. Dette kan enten gjøres manuelt (noe jeg vil anbefale hvis du allikevel skal skrive inn dataene) eller ved hjelp av CONVert-kommandoen.

CONVert-kommandoen koder om en kolonne fra tall til tekst (eller omvendt) ut fra kolonner som inneholder en konverteringsnøkkel. Den generelle formen er

```
CONVert nøkkel1 nøkkel2 kildekolonne målkolonne
```

hvor *nøkkel1* er en kolonne som inneholder en liste med tall, *nøkkel2* er en kolonne med tekstverdiene som tilsvarer tallene i *nøkkel1*, *kildekolonne* inneholder tallene som skal konverteres, og *målkolonne* er kolonnen hvor teksten skal inn. Nøkkelkolonnene må stå i samme rekkefølge som konverteringen skal skje — dvs. konverterer du fra tall til tekst må *nøkkel1* inneholde teksten, *nøkkel2* tallene. Et eksempel:

```
MTB > print c1-c3
```

ROW	C1	C2	C3	(tekstvariablene lagt inn ved hjelp av dataeditoren)
1	1	Mann	2	


```
2    2    Kvi nne    2
3
4
5
```

```
MTB > convert c1 c2 c3 c4 (C4 blir C3 som tekstkolonne
MTB > print c1 - c4      med nøkler i C1 og C2)
```

ROW	C1	C2	C3	C4
1	1	Mann	2	Kvi nne
2	2	Kvi nne	2	Kvi nne
3			1	Mann
4			1	Mann
5			2	Kvi nne

```
MTB > convert c1 c2 c4 c5 (au da — feil rekkefølge)
* ERROR * Mismatched data and conversion table
```

Som synes skriver Minitab ut en feilmelding hvis du kommer til å bytte om rekkefølgen.

Manipulering med kolonner

I dette kapitlet tar vi for oss endel kommandoer som benyttes til manipulering med kolonner — dvs. slike ting som kopiering, "oppklipping", sletting, omkodning osv. Hvis du har Minitab versjon 7.1 eller nyere for en IBM-kompatibel PC, eller versjon 8 eller nyere for en Macintosh, kan du i mange tilfelle benytte et eget program (fullskjerm dataeditor) for å gjøre mange av de tingene som er nevnt som egne kommandoer her. Dette programmet er særlig praktisk når du skal forandre på enkelt-tall i en kolonne, eller andre "mindre" forandringer.

4.5 Å gi kolonner navn

Kolonner, matriser og konstanter har som før nevnt betegnelsene C, M og K, med et tall etter. Dette er i og for seg greit nok — men når du etterhvert har laget deg endel kolonner kan det bli vanskelig å huske hvilke kolonner som inneholder hvilke data.

Da er NAME-kommandoen grei å bruke. Sett at du har en kolonne C3 som inneholder alderen til endel personer. Du kan da gi kommandoen

```
NAME C3 'Alder'
```

Dermed vil du kunne bruke betegnelsen 'Alder' (apostrofene må være med) i stedet for C3 i alle kommandoer du ønsker å gi som vedrører denne kolonnen. (Du kan fremdeles benytte C3 hvis du ønsker det.)

Du kan godt navngi flere kolonner samtidig, men hver kolonne må ha et entydig navn. Dette vil si at

```
NAME C5 'Pris Kr' C6 'Vekt'
```

er tillatt, mens

```
NAME C5-C7 'Resultat'
```

ikke er det.

Kolonnenavnet (eller navnet på konstanten eller matrisen) kan ha inntil 8 bokstaver. Alle tegn er tillatt, med følgende unntak:

- navn kan ikke begynne eller slutte med mellomrom
- apostrof (') og nummertegn (#) kan ikke brukes
- både store og små bokstaver kan benyttes. Minitab oppfatter imidlertid ikke forskjellen, f.eks. blir 'Vekt' og 'VEKT' oppfattet av Minitab som den samme kolonnen.
- "Norske" tegn (dvs. æ, ø, å, Æ, Ø, Å) kan brukes, men vil på skjermen se ut som et mellomrom. Det enkleste er etter min mening å bruke "e" eller "ae" i stedet for "æ", "o" eller "oe" for "ø" og "a" eller "aa" for "å" (og dessuten skrive til Norsk Språkråd og foreslå at vi kutter ut disse bokstavene fra alfabetet — siden de bare skaper problemer på de fleste datamaskiner. . .).

4.6 Å kopiere kolonner

For å kopiere en eller flere kolonner brukes COPY-kommandoen. Denne kommandoen kommer du til å bruke mye — samtidig som den illustrerer mange sider av det å gi kommandoer i Minitab svært godt. Kommandoen er særlig egnet til å lage subset (delmengder) av data.

La oss begynne med et enkelt eksempel: Du skal kopiere kolonne C7 til en ny kolonne. Du velger å ha kopien av C7 i kolonne C19. Du vil da gi følgende kommando:

```
MTB > COPY C7 C27
MTB > _
```

Kolonne C7 er nå kopiert til kolonne C29 — dvs. at du har to kolonner som er like. Det er en liten forskjell på dem; hvis du har gitt kolonne 7 et navn ved hjelp av NAME-kommandoen, vil ikke den nye kolonnen (C27) få med seg dette.

Det går også an å kopiere flere kolonner samtidig, f.eks. slik:

```
MTB > COPY C7 C20 C30 C31
MTB > _
```

I dette tilfellet kopieres kolonne 7 til kolonne 30 og kolonne 20 til kolonne 31.

Du kan kopiere så mange kolonner du vil — forutsatt at du har et like antall kolonner (dvs. like mange kildekolonner som målkolonner). Kolonnene kan spesifiseres slik:

C1, C3	betyr	kolonne 1 og 13
C1-C3	betyr	kolonne 1, 2 og 3

hvis du spesifiserer en tom kolonne som kildekolonne, avbrytes **hele** kopieringskommandoen med denne feilmeldingen:

```
ERROR * NO DATA IN COLUMN
```

COPY-kommandoen har to viktige subkommandoer: USE og OMIT (se neste avsnitt).

På samme måte som med kolonner kan du kopiere konstanter og matriser. Det er også mulig å kopiere mellom disse dataelementene (f.eks. fra en kolonne til en rekke med konstanter) uten at det skal gjennomgås her. For mer detaljert informasjon om dette: gi Minitab kommandoen HELP COPY.

4.6.1 Mer spesialisert kopiering

Kopiering av hele kolonner er greit nok — men hva hvis du kun skal ha med deg *noen* av radene i en kolonne over i en annen?

Du har to subkommandoer til COPY til disposisjon: USE og OMIT. Med USE spesifiserer du hvilke rader du skal ha med deg, og med OMIT spesifiserer du hvilke rader du *ikke* skal ha med.

Radene kan spesifiseres på to måter; enten ved radnummer, eller ved at du spesifiserer kriterier for å kopiere dem.

Forvirret? La oss se på noen eksempler!

Du skal kopiere f.o.m. rad 10 t.o.m. rad 20 fra kolonne C2-C5 (som hver inneholder 20 rader) inn i kolonnene C10-C13. Du kan gjøre dette på to måter, enten med USE:

```
COPY C2-C5 C10-C13;  
USE 10: 20.
```

eller med OMIT:

```
COPY C2-C5 C10-C13;  
OMIT 1: 9.
```

Vi kan også bruke USE eller OMIT til å spesifisere hvilke rader som skal kopieres ut fra hva de inneholder:

Du har tre kolonner: C1 inneholder kjønn (1 = mann, 2 = kvinne), C2 kroppshøyde og C3 kroppsvekt på endel personer. Du ønsker å splitte opp disse kolonnene slik at mennenes høyde og vekt ligger i henholdsvis C10 og C11, og kvinnenenes i C20 og C21. Dette kan gjøres på ulike måter, enten med USE:

```
COPY C2 C3 C10 C11;  
USE C1 = 1.
```

Dette kopierte alle rader fra kolonne 2 og 3, hvor den tilsvarende raden i kolonne 1 var lik 1, inn i kolonne 10 og 11. Eller, med andre ord: høyde og vekt for alle mennene i utvalget ligger nå i kolonne 10 og 11.

For forandringens skyld kan vi benytte OMIT når vi skal kopiere dataene for kvinnene:

```
COPY C2 C3 C20 C21;  
OMIT C1 = 1.
```

Dette kopierte alle rader fra kolonne 2 og 3, hvor den tilsvarende raden i kolonne 1 **ikke** var lik 1, inn i kolonne 10 og 11. *Merk deg at dette inkluderer rader som inneholder alle andre tegn enn 2 — f.eks. * (manglende observasjon).*

4.7 Å slette data eller kolonner

Kommandoen for å slette en eller flere kolonner er

ERASe kolonner

f.eks. *ERASE C1-C4, C6*. Den kan også benyttes på konstanter og matriser.

Hvis du skal slette rader fra en kolonne (men ikke hele kolonnen), bruker du *DELEte*-kommandoen. Den gis slik:

DELEte rad(er) kolonne(r)

For å slette rad 3 t.o.m. 8 fra kolonne C6 t.o.m. kolonne C10 gir du denne kommandoen:

```
MTB > DELEte 3:8 C6-C10  
MTB > _
```

Hvis du gir DELEte-kommandoen uten å spesifisere rader, får du en feilmelding. For å slette hele kolonner, bruk ERASe-kommandoen.

Hvis du ønsker å slette *alle* kolonner, konstanter og matriser i en jafs, kan du benytte kommandoen

REStart

Den sletter alle data du har tilgjengelig — dvs. at den tilsvarer en STOP og ny oppstart av Minitab (noe navnet indikerer). Dette vil også si at alle kommandoer du tidligere har gitt (f.eks. OUTFile, OH el.l.) mister sin effekt.

4.8 Å slå sammen flere kolonner til én

Hvis du skal slå sammen to kolonner til én (dvs. "stable dem oppå hverandre") kan du bruke STACK-kommandoen. La oss si at du skal "skjømte sammen" kolonne 1 og 2 (med kolonne 1 øverst) og lagre dem i kolonne 20. Du ville da gi denne kommandoen:

```
MTB > STACK C1 C2 C20  
MTB > _
```

Du kan skrive hvor mange kolonner du vil — de blir alle sammen slått sammen i den rekkefølgen du har skrevet og lagt inn i den siste kolonnen.

Du kan også bruke kommandoen til å slå sammen mange "kildekolonner" inn i to eller flere "målkolonner". Dette gjør du ved hjelp av parenteser rundt hver gruppe kolonner. For eksempel

```
MTB > STACK (C1 C2) (C3 C4) C5 C6
```

stabler C1 oppå C3 og putter resultatet i C5, C2 oppå C4 og putter resultatet i C6. Det er viktig at du har like mange kolonner innenfor parentesene som du har "målkolonner" (i dette tilfelle to).

4.9 Å dele opp kolonner

Hvis du skal dele en kolonne opp i flere kolonner, kan du benytte UNSTACK-kommandoen. Denne kommandoen vil du nesten alltid benytte subkommandoen SUBScripts til.

Ta følgende eksempel: Vi har en kolonne (C3) som inneholder vekten til endel personer. I en annen kolonne (C1) har vi disse personenes kjønn (1 = mann, 2 = kvinne). Vi ønsker å dele opp kolonnen C3 slik at vi får to kolonner: C10 og C11, som inneholder henholdsvis vekten for menn og vekten for kvinner. Dette gjør vi på denne måten:

```
MTB > UNSTack C3 C10 C11;
```

```
SUBC> SUBScri pts C1.  
MTB > INFO
```

COLUMN	NAME	COUNT	MISSING
C1	Kj onn	216	
C3	Vekt	216	4
C10		154	
C11		62	4

CONSTANTS USED: NONE

(Info-kommandoen til slutt er ikke nødvendig — den er med for illustrasjonens skyld). Det er mulig å bruke UNSTack på flere kolonner samtidig — omtrent som STACK-kommandoen.

4.10 Å forandre data i en kolonne

Av og til har du bruk for å forandre dataene i en kolonne. Dette kan gjøres med LET-kommandoen (s.d.), men en mer elegant måte er som regel å benytte CODE-kommandoen. Denne kommandoen fungerer omtrent slik en forandringskommando i en tekstbehandlingspakke fungerer; den forandrer alle data av en viss type til det du oppgir.

Denne kommandoen gis slik:

CODE (x) y kildekolonne målkolonne

hvor x er den verdien som skal kodes om (parentesene må alltid være med), y er den verdien det skal kodes til, *kildekolonne* er den kolonnen de opprinnelige dataene ligger, og *målkolonne* er den kolonnen dataene skal lagres i.

Du kan kode om flere verdier, eller rekker av verdier, ved å spesifisere dem innenfor parentesene f.eks. slik:

```
(1 3 4) gir tallene 1 3 4  
(1: 4) gir tallene 1 2 3 4
```

Du kan også spesifisere flere omkodinger samtidig.

Forvirret? La oss ta følgende eksempel: I en kolonne C3 har du liggende alderen for en rekke mennesker. Sett at du ønsker å kode om alderen (som er oppgitt i hele år) til en kode for "under tredve" og en for "tredve og eldre".

Dette kan da gjøres slik:

```
CODE (0: 29) 1 (30: 100) 2 C3 C4
```

I kolonne 4 vil det nå ligge 1-tall og 2-tall, som representerer henholdsvis "under tredve" og "tredve og eldre".

Omkodinger er særlig viktig ved utforming av spørreskjemaer — ved å bruke omkoding får du selv muligheten til å kategorisere data. Hvis du spør folk om de røyker mye, vil du kunne få svært upresise svar — fordi meningen av hva som er "mye" røyking varierer fra person til person. Hvis du i stedet spør "hvor mange sigaretter røyker du om dagen/uken", vil du selv kunne bestemme hvor grensen for "mye" og "lite" går — og kunne dele respondentene inn vha. CODE-kommandoen.

Ved å gruppere data på denne måten gjør du ofte en statistisk analyse lettere; ofte er det mer interessant å se på grupperinger enn de enkelte tilfellene. Ofte brukes dette for å kunne foreta analyser som forutsetter diskrete data i stedet for kontinuerlige — f.eks. ved å gruppere oppgaver over inntekter eller kostnader i grupper (høyt- og lavtlønte el.l.).

4.11 Å liste opp innholdet i en kolonne

Hvis du ønsker å se hvilke data en kolonne inneholder, kan du gi kommandoen

```
PRINt kolonne(r)
```

Denne kommandoen skriver ut innholdet av kolonnene *på skjermen* (altså ikke på skriver — da må du bruke PAPER eller OUTFILE-kommandoen).

Hvis du ber om utskrift av en enkeltkolonne, blir den skrevet ut i hele skjermens bredde (dvs. at tallene ikke kommer under hverandre). Hvis du spesifiserer to eller flere kolonner, blir kolonnene listet opp ved siden av hverandre, rad for rad.

4.12 Å endre enkelt-tall (LET-kommandoen)

Kommandoen gis slik:

```
LET variabel = uttrykk
```

"Variabel" kan være en kolonne, en konstant eller en matrise.

Eksempel: `LET C1 = C2 + C3 / 12`
 `LET 'Moms' = 'Pri s' * .2`

Rekkefølgen for utregninger er slik:

Først funksjoner (kommandoer) deretter potenser, deretter multiplikasjon/divisjon, deretter addisjon/subtraksjon. Uttrykkene regnes fra venstre mot høyre, og uttrykk innenfor parenteser regnes ut før parentesene løses opp.

Operatorer (regnetegn) er følgende:

+	addisjon
-	subtraksjon
*	multiplikasjon
/	divisjon
**	potens ("opphøyet i"). NB! Må være ** ikke *!

Det vil altså si at

```
LET K1 = 5 + 5 / 5     gir K1 = 6  
LET K1 = (5 + 5) / 5   gir K1 = 2
```

LET-kommandoen kan også benyttes til å forandre en enkelt verdi i en kolonne — noe som er greit hvis du skal endre en enkelt observasjon f.eks. som følge av punchefeil. Dette gjøres på denne måten:

LET kolonne(rad) = uttrykk

f.eks.

LET C4(26) = 75.67

som setter den 26. observasjonen (raden) i kolonne 4 til 75.67. (På en PC eller Macintosh vil du som regel benytte dataeditoren til dette).

Det finnes også sammenlignende og logiske operatører til LET-kommandoen. Disse er

=	eller	<i>EQ</i>	:	er lik (equal to)
~=	eller	<i>NE</i>	:	er ikke lik (not equal to)
<	eller	<i>LT</i>	:	mindre enn (less than)
>	eller	<i>GT</i>	:	større enn (greater than)
<=	eller	<i>LE</i>	:	mindre enn eller lik (less than or equal to)
>=	eller	<i>GE</i>	:	større enn eller lik (greater than or equal to)
&	eller	<i>AND</i>	:	og samtidig
/	eller	<i>OR</i>	:	eller
~	eller	<i>NOT</i>	:	ikke (negasjonstegn)

Disse sammenligningene og sammensetningene foretas radvis. Hvis vilkåret er sant for den enkelte rad, vil det bli satt et ettall i resultatkolonnen, hvis ikke vil det bli satt en null. F.eks. vil setningen

LET C11 = (C1 GE 5) AND NOT (C2 = 1)

gjøre at det i C11 vil være ettall for alle rader hvor C1 er større enn eller lik 5 og samtidig C2 ikke er lik 1. Se punkt ?, side 59 for et eksempel på hvordan dette kan benyttes.

4.12.1 Funksjoner til LET-kommandoen

Det finnes en mengde (matematiske, statistiske og andre) funksjoner til LET-kommandoen. Alle disse funksjonene finnes også som egne kommandoer (se bl.a. kapittelet om beskrivende kommandoer) uten at jeg vil komme nærmere inn på dem her.

For alle funksjonenes vedkommende gjelder denne formen:

LET kolonne/konstant/matrise = funksjon(argument)

for eksempel

LET K2 = SUM(C7)

dvs. "La konstant nr. 2 være lik summen av alle cellene i kolonne 7", eller

$LET K2 = SIN(0.80)$

Her følger en liste over funksjonene med (tildels svært) kortfattede forklaringer:

Matematiske funksjoner

ABSOLute	Absoluttverdien $SQRT(x^{**2})$	(tilsvarer	<p>Minitab har ingen INTEGER-kommando, det vil si en kommando som tar vekk desimaldelen av et tall, slik at bare heltallet gjenstår. I statistikk kan du ha bruk for dette hvis du har kompliserte kategoridata, f.eks. av samme type som USA's SIC (Standard Industry Code). Hvis du har firesifrede SIC-koder for en rekke firma, og ønsker å redusere dem til tosifrede, kan du gjøre dette med kommandoen</p> <p style="text-align: center;">$LET C2 = ROUND ((C1/100) - 0.5)$</p> <p>i stedet for å bruke RECOde-kommandoen. Legg merke til at dette ikke fungerer hvis du har negative data.</p>
SIGNs	Fortegnet til argumentet. Gir 1 hvis argumentet er større enn null (positivt), 0 hvis argumentet er 0, og -1 hvis argumentet er mindre enn 0 (negativt).		
ROUND	Gir avrundet verdi av argumentet (avrundet tilbake til desimalpunktet). 2.21 avrundes til 2, 3.50 avrundes til 4.		
SQRT	Gir kvadratrotten av argumentet. Hvis argumentet er negativt, kommer det en feilmelding, og resultatet settes til "*" ("manglende verdi").		
LOGE	Gir den naturlige logaritme, dvs. det tall e (2.731828128459045. .) må opphøyes i for å gi argumentet. Hvis argumentet er 0 eller negativt, kommer det en feilmelding, og resultatet settes til "*" ("manglende verdi").		
EXPOntiate	Gir den inverse naturlige logaritme, dvs. opphøyer e i argumentet.		
ANTIlog	Opphøyer 10 i argumentet. (ANTILOG 3 = 1000)		
PARSums	Kumulerer tallene i en kolonne. Dette vil si at det lages en ny kolonne hvor hver rad er summen av de ovenforliggende radene i utgangskolonnen.		
	Eksempel:		
	Kolonne 1	Kolonne 2 (PARSums)	
	10	10	
	2	12	
	14	26	
	11	37 osv. . . .	
PARProducts	gir det samme som PARSUMS — bortsett fra at hver rad i kolonne 2 er <i>produktet</i> av de ovenforliggende radene i kolonne 1.		
	Eksempel:		

Kolonne 1	Kolonne 2 (PARProducts)
10	10
2	20
14	280
11	3080 osv. . . .

RANK Rangering av innholdet i en kolonne (denne funksjonen er forklart i punkt ?).

SORT Sortert utgave av en kolonne (denne funksjonen er forklart i punkt ?).

LAG forskyver kolonner. Denne funksjonen/kommandoen benyttes når du ønsker å flytte radene i en kolonne "nedover" til en annen kolonne. LAG er nyttig ved tidsserieanalyser, hvor du ser på i hvilken grad verdiene i en periode er en funksjon av perioden foran.

Kommandoen gis slik:

LAG n Cx Cy

hvor *n* er antall rader kolonne *x* skal "forsinkes". (Hvis du ikke spesifiserer noe, er *n* = 1). Et eksempel:

LAG 2 C7 C14

flytter dataene i kolonne 7 2 rader "nedover", og lagrer resultatet i kolonne 14.

Merk at de to øverste radene i kolonnen til høyre er tomme (dvs. de inneholder "*"), samt at den nye kolonnen er like lang som den opprinnelige. Dette vil si at de to nederste radene i utgangskolonnen (C7 i eksemplet ovenfor) ikke er med i målkolonnen (C14).

LAG kan også benyttes til å lage "rullerende gjennomsnitt". Hvis du har månedlige data liggende i kolonne C1, vil denne kommandosekvensen lage 3-måneders rullerende gjennomsnitt i kolonne C14:

```
MTB > LAG C1 C11
MTB > LAG 2 C1 C12
MTB > LAG 3 C1 C13
```

$$MTB > LET C14 = (C11 + C12 + C13)/3$$

(Men husk at resultatet ikke vil stemme for de to første radene).

Statistiske funksjoner

Alle disse funksjonene er beskrevet nærmere i kapittelet om "enkle, deskriptive kommandoer" (kapittel ?, side 41).

COUNT	antall dataelementer i en kolonne, dvs. N + NMISs
N	antall observasjoner i en kolonne (dvs. COUNT – NMISs)
NMISs	antall "missing values" i en kolonne (dvs. antall "*")
SUM	summen av alle tallene i en kolonne
MEAN	det aritmetiske (dvs. "vanlige") gjennomsnittet av en kolonne
STDEV	standardavviket til en kolonne
MEDIAN	medianen til tallene i en kolonne
MINIMUM	det minste tallet i kolonnen
MAXIMUM	det største tallet i kolonnen
SSQ	kvadratsummen av kolonnen

Geometriske funksjoner

SIN	Sinus
COS	Cosinus
TAN	Tangens
ASIN	Arcsinus
ACOS	Arccosinus
ATAN	Arctangens

For alle disse funksjonene gjelder at vinkler oppgis i radianer. Formelen for konvertering er

$$\text{grader} = \text{radianer} * 57.297$$

$$\text{radianer} = \text{grader} / 0.017453$$

Gal bruk av funksjonene (som f.eks. å spørre etter Arcsinus til 1.2) vil resultere i en feilmelding, samt at resultatet av funksjonen blir "*" (dvs. koden for "manglende verdi").

4.13 Å sortere en kolonne

Under dette punktet finner vi to kommandoer: SORT og RANK. SORT sorterer en eller flere kolonner. RANK rangerer en kolonne, dvs. finner ut av rekkefølgen i den — men flytter ikke på noen data.

4.13.1 Sortering

Kommandoen for å sortere en eller flere kolonner er

SORT kolonne(r) kildekolonne(r)

Du må oppgi hvilke(n) kolonne(r) som skal sorteres — og hvor de sorterte data skal puttes inn (tilsvarende COPY-kommandoen). Hvis du sorterer flere kolonner, blir de sortert etter innholdet i den **første** kolonnen. La oss ta et eksempel:

```
MTB > SORT C1 C21
MTB > _
```

Med andre ord: Kolonne 21 inneholder nå dataene fra kolonne 1, men sortert i stigende rekkefølge.

```
MTB > SORT C1 C2 C3 C11 C12 C13
MTB > _
```

Her blir kolonne 1 sortert — og i kolonne 2 og 3 blir dataene lagt i rekkefølgen til kolonne 1. Resultatet av sorteringen blir puttet inn i kolonnene 11-13. På samme måte som COPY-kommandoen må det alltid være like mange "kilde-" som "målkolonne" — og du kan godt la "målkolonnene" være de samme som "kildekolonnene" (men da mister du de opprinnelige, usorterte dataene). Ved like verdier (dvs. to like verdier i en kolonne) blir de i den sorterte kolonnen liggende i samme rekkefølge som de lå i i kildekolonnen³. "Manglende observasjon" (*) blir liggende sist.

Alfanumeriske variable blir sortert alfabetisk. Hvis samme bokstav, går små bokstaver (minuskler) før store (versaler). Ved sortering av alfanumeriske kolonner blir "manglende observasjon" (som markeres med et mellomrom) liggende først i den ferdigsorterte kolonnen.

SORT har to subkommandoer (f.o.m. versjon 7.1); BY og DESCending. Hvis du gir subkommandoen DESCending, blir dataene i kolonnen(e) sortert i fallende rekkefølge — dvs. med de største tallverdiene først. Subkommandoen BY lar deg mer detaljert spesifisere hvilke kolonner som skal danne grunnlag for sorteringen. Et eksempel:

```
MTB > SORT C1 C2 C21 C22;
SUBC> BY C4 C5.
```

I dette tilfelle sorteres kolonne C1 og C2 etter rekkefølgen i kolonne C4 og C5 (dvs. hvis det er like data i C4, sorteres det etter C5). Resultatet puttes inn i C21 og C22. *NB!* C1 har i dette tilfelle ingen innflytelse på rekkefølgen i sorteringen — når du bruker BY-subkommandoen er det kun den som gjelder (med andre ord, hvis du vil ha dataene sortert på grunnlag av C1, må du oppgi C1 først etter BY).

4.13.2 Rangering (sortering uten flytting)

Å rangere data vil si å sortere dem — men uten at de blir flyttet. Et eksempel på rangering finner vi i resultatlister etter store idrettsstevner (Oslo Maraton o.l.) — disse publiseres gjerne sortert alfabetisk, med et tall etter navnet på hver enkelt løper som forteller hvilken plassering løperen fikk.

³Dette gjelder f.o.m. utgave 7.1 av Minitab. I de tidligere utgavene er det en feil som gjør at disse verdiene ikke alltid blir liggende i den samme rekkefølge som før sorteringen.

Kommandoen for å rangere en kolonne er

`RANK kildekolonne målkolonne`

Hvor "kildekolonne" er den kolonnen som skal rangeres (som inneholder resultatene, om du vil) og "målkolonne" er en kolonne som etter at rangeringen er gjennomført vil inneholde "plasseringene" til dataene i kildekolonnen. Målkolonnen vil altså inneholde tall fra 1 (som representerer laveste verdi i kildekolonnen) oppover til antall rader i kildekolonnen.

Hvis to data i kildekolonnen er like, blir verdien i målkolonnen et "gjennomsnitt" av plasseringen. Dette vil si at to tall som "deler 3. plass" får verdien 3.5 — gjennomsnittet av 3. og 4. plass (se tekstboks).

```
MTB > RANK C1 C2
MTB > PRINT C1 C2
```

ROW	C1	C2
1	34	2.0
2	56	5.0
3	16	1.0
4	38	3.5
5	38	3.5
6	70	6.0

Enkle deskriptive kommandoer

I dette kapitlet skal vi ta for oss kommandoer som beskriver egenskaper til data ved hjelp av endel observatorer (nøkkeltall). Disse kommandoene er greie å bruke når du skal skaffe deg en oversikt over hvordan et datasett ser ut.

4.14 DESCribe (beskriv)

DESCribe er en meget nyttig kommando, fordi den inneholder en rekke beskrivende elementer samtidig. Et eksempel: Forestill deg at du har endel tall i kolonne 1 og gir kommandoen

```
DESCribe C1
```

Minitab svarer f.eks:

	<i>N</i>	<i>N*</i>	<i>MEAN</i>	<i>MEDI AN</i>	<i>TRMEAN</i>	<i>STDEV</i>
<i>C1</i>	560	40	10. 429	10. 500	10. 476	4. 156
	<i>SEMEAN</i>	<i>MI N</i>	<i>MAX</i>	<i>Q1</i>	<i>Q3</i>	
<i>C1</i>	0. 176	3. 000	17. 000	7. 000	14. 000	

Her betyr

<i>N</i>	Antallet observasjoner (dvs. antall rader i kolonnene (COUNT) minus antall "missing value").
<i>N*</i>	Antallet "missing values" (dvs. "ikke observert").
<i>MEAN</i>	Det aritmetiske ("vanlige") gjennomsnittet.
<i>MEDI AN</i>	Medianen (den "midterste" observasjonen).
<i>TRMEAN</i>	"Trimmed mean" (justert gjennomsnitt) — beregnet ved at Minitab har "klippet vekk" de høyeste og laveste 5 % av verdiene i kolonnen — og beregnet gjennomsnittet av resten. Dette er en enkel metode for å justere for unormalt høye og lave observasjoner (ekstremverdier, outliers).
<i>STDEV</i>	Kolonnens standardavvik.
<i>SEMEAN</i>	Gjennomsnittets standardavvik (standard error of the mean). Tilsvare standardavviket dividert med kvadratroten av antall observasjoner, dvs. $STDEV/\sqrt{N}$.
<i>MAX</i>	Maksimum; kolonnens høyeste verdi.
<i>MI N</i>	Minimum; kolonnens laveste verdi.
<i>Q1</i>	Nedre kvartil.
<i>Q3</i>	Øvre kvartil.

DESCribe har en subkommando; BY. Denne kommandoen lar deg spesifisere en "kategorikolonne", og produserer den samme informasjon som DESCribe, men oppdelt for hver kategori i "kategoriseringskolonnen". "Kategoriseringskolonnen" kan kun inneholde heltall (integers) eller "*".

Innviklet? La oss ta et eksempel:

Du har målt intelligenskvotienten på endel svensker og nordmenn, puttet IQ-tallene i kolonne

C1, og nasjonaliteten (1 = Nordmann, 2 = Svenske) i kolonne C2. Hvis du gir kommandoen

```
DESCRibe C1;  
BY C2.
```

kommer denne informasjonen (eller noe lignende) frem:

	C2	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
C1	1	7	117.43	100.00	117.43	24.22	9.15
	2	12	104.00	100.00	101.50	16.62	4.80

	C2	MIN	MAX	Q1	Q3
C1	1	99.00	160.00	100.00	137.00
	2	88.00	145.00	90.75	111.00

(Hvis det ovenstående skulle tyde på at svensker er mindre intelligente enn nordmenn, kan du jo trekke frem utvalgsskjevhet. . . .)⁴.

4.15 Andre beskrivende kommandoer

(De fleste av kommandoene under denne headingen er inkludert i DESCRibe-kommandoen — men av og til har du ikke bruk for alle de opplysningene DESCRibe gir).

Alle beskrivende kommandoer gis på samme måte, f.eks slik:

```
MTB > COUNT C3  
COUNT = 156  
MTB > _
```

Du kan også gi beskjed om at resultatet av kommandoen (som alltid er et tall) skal lagres i en konstant; f.eks. kan du gi kommandoen

```
MEAN C3 K4
```

som tar gjennomsnittet av tallene i kolonne 3 og lagrer det i konstanten K4.

Her er en liste over slike typiske beskrivende kommandoer.

<u>Kommando</u>	<u>Beskrivelse</u>
COUNT	antall dataelementer i en kolonne, dvs. N + NMISs
N	antall observasjoner i en kolonne (dvs. COUNT – NMISs)
NMISs	antall "missing values" i en kolonne

⁴Lugna ner sig, nu — det här är *fiktiva* observationer

SUM	summen av alle tallene i en kolonne
MEAN	det aritmetiske (dvs. "vanlige") gjennomsnittet av en kolonne
STDEv	standardavviket til en kolonne
MEDIan	medianen til tallene i en kolonne. Dette tallet fremkommer ved å sortere tallene, og så finne det tallet som er nummer $(n+1)/2$ i rekken.
MINImum	Det minste tallet i kolonnen
MAXImum	Det største tallet i kolonnen
SSQ	Kvadratsummen av kolonnen, dvs. summen av alle elementene etter at de er opphøyet i annen potens (dvs. ganget med seg selv). SSQ C3 er det samme som $SUM(C3**2)$.

TABLE-kommandoen

TABLE-kommandoen er kanskje den mest brukte kommandoen i Minitab — i hvert fall innenfor deskriptiv analyse, dvs. når du kun skal beskrive et datasett. TABLE kan brukes på en enkelt kolonne — til å gi en kjapp oversikt over fordeling av variable — eller den kan brukes på to eller flere kolonner — til *krysstabulering*. TABLE-kommandoen har en rekke subkommandoer som kan gi deg mye informasjon på en kjapp og elegant måte.

I de følgende eksempler vil jeg benytte et datasett med disse kolonnene:

```
MTB > INFOrmation
```

<i>COLUMN</i>	<i>NAME</i>	<i>COUNT</i>	<i>MISSING</i>
<i>C1</i>	<i>Kjonn</i>	<i>177</i>	
<i>C2</i>	<i>Hoyde</i>	<i>177</i>	
<i>C3</i>	<i>Vekt</i>	<i>177</i>	
<i>C4</i>	<i>Skonr.</i>	<i>177</i>	
<i>C5</i>	<i>Al der</i>	<i>177</i>	
<i>C6</i>	<i>Sigaret t</i>	<i>177</i>	<i>4</i>
<i>C7</i>	<i>Royki ng</i>	<i>177</i>	<i>4</i>

Disse dataene er noen av resultatene fra et spørreskjema som ble delt ut blant Siviløkonomstudentene (83-kullet) på BI. Kolonne 6 (Sigarett) inneholder antall sigaretter en person røyker pr. uke. Kolonne 7 (Royking) som inneholder 1-tall hvis personen røyker og 0 hvis han/hun ikke røyker. En fullstendig utlisting av disse dataene finner du bakerst i manualen.

Til å begynne med tar vi for oss kolonne 1 (Kjønn). Vi gir denne kommandoen:

```
MTB > TABLE C1
```

```
ROWS: Kjonn
```

```
COUNT
```

<i>1</i>	<i>145</i>
<i>2</i>	<i>32</i>
<i>ALL</i>	<i>177</i>

Eller, med andre ord, det er 145 menn og 32 kvinner i utvalget. . . .

La oss så se på to kolonner samtidig:

```
MTB > table c5 c1
```

	COLUMNS: Kj onn		
ROWS: Al der	1	2	ALL
19	4	1	5
20	28	14	42
21	51	8	59
22	39	3	42
23	10	3	13
24	8	1	9
25	4	1	5
26	0	1	1
36	1	0	1
ALL	145	32	177

CELL CONTENTS --
COUNT

Her får vi litt mer informasjon: Vi kan f.eks. se at de fleste personene er 21 år gamle — og at av disse er 51 menn og 8 kvinner. Den observante leser har forlenget lagt merke til at typetallet (det oftest forekommende tallet) for kvinnene er 1 år lavere enn den for mennene — militærtjenesten har muligens skylden for dette?

Legg merke til at den kolonnen som kommer først i TABLE-kommandoen blir *radene* i tabellen (heretter kalt tabellradene), den neste blir *kolonnene* (heretter kalt tabellkolonnene). Hvert "kryss" av tabellrader og -kolonner kalles en celle. Det er mulig å tabulere flere enn to kolonner mot hverandre — da blir det en tabell for hver forekomst i den tredje kolonnen — men dette er uoversiktlig og sjelden nødvendig, da man som regel kan benytte subkommandoer (se neste pkt.) for å vise statistikk om andre kolonner enn de man er interessert i.

5.1 Sub-kommandoer til TABLE

TABLE-kommandoen har en rekke subkommandoer. De brukes på samme måte som subkommandoer benyttes overalt i Minitab; ved at de puttes etter et semikolon på linjer under hovedkommandoen. Det er først og fremst subkommandoene som gjør TABLE-kommandoen interessant — ved hjelp av dem kan du fremstille tildels svært komplekse sammenhenger på en oversiktlig måte.

Stadig vekk med utgangspunkt i datasettet beskrevet ovenfor; ta en titt på dette eksemplet:

```
MTB > TABLE c1 c7;
SUBC> MEANS c2;
SUBC> MEANS c3.
```

ROWS: Kj onn	COLUMNS: Royki ng		
	0	1	ALL

1	182.30 73.472	182.55 75.364	182.35 73.915
2	169.79 55.658	170.54 60.231	170.09 57.516
ALL	180.43 70.807	179.15 71.087	180.09 70.882

CELL CONTENTS --

Høyde: MEAN

Vekt: MEAN

Her benyttes subkommandoen MEANs — og den benyttes med andre kolonner enn de to som blir "tablett" mot hverandre. Dette er en meget effektiv måte å benytte TABLE på — den gir masse informasjon på en rask måte, uten for mye "ekstra" tall.

Av utskriften fremgår det f.eks. at gjennomsnittshøyden for kvinner som røyker er 170,54 cm — og for de som ikke røyker er den 169,79. Gjennomsnittshøyden for **alle** kvinnene i utvalget er 170,09 cm (som du ser under overskriften "ALL").

Nedfor følger en liste av de subkommandoer du kan benytte til TABLE-kommandoen — med kortfattede forklaringer. Endel av disse kommandoene er forklart nærmere under kapittelet om "enkle, deskriptive kommandoer" (kapittel ?):

MEANs	Gjennomsnitt
MEDIans	Median
SUMS	Sum
MINImums	Laveste verdi
MAXImums	Høyeste verdi
STDEv	Standardavvik
N	Antall observerte med resultat
NMISsing	Antall "manglende verdi"
COUNts	N + NMISsing, dvs. antall rader
STATs	gir antall observerte (N), gjennomsnitt (MEANs) og standardavvik (STDEv). Eksempel:

```
MTB > TABLE c7 c1;
SUBC> STATs c2.
```

```
ROWS: Royki ng      COLUMNS: Kj onn
           1          2          ALL
0         108         19         127
          182.30     169.79     180.43
```

	6.54	5.18	7.76
1	33	13	46
	182.55	170.54	179.15
	6.06	4.77	7.88
ALL	141	32	173
	182.35	170.09	180.09
	6.41	4.95	7.79

CELL CONTENTS --
 Hoyde: N
 MEAN
 STD DEV

DATA Lister opp dataene innenfor hver celle. Bør ikke brukes ved store datamengder.

PROPORTION gir andelen av data som passer til de kriterier du spesifiserer. For eksempel:

```
MTB > TABLE 'Kjonn' 'Røyking';
SUBC> PROPORTION 0 40 'Skonr.'.
```

ROWS: Kjonn COLUMNS: Røyking

	0	1	ALL
1	0.0168	0.0909	0.0329
2	0.9286	0.9500	0.9355
ALL	0.2547	0.4151	0.2944

CELL CONTENTS --
 Skonr.: PROPORTION IN RANGE 0 TO 40

Her vises andelen kvinner/menn og røykere/ikke-røykere som har skonummer 40 eller mindre.

ROWPercents gir den den andel (i prosent) cellen utgjør av tabellraden
 COLPercents gir den den andel (i prosent) cellen utgjør av tabellkolonnen
 TOTPercents gir den den andel (i prosent) cellen utgjør av hele tabellen

Eksempel på disse subkommandoene:

```
MTB > TABLE c7 c1;
SUBC> ROWPercent;
SUBC> COLPercent;
SUBC> TOTPercent.
```

ROWS: Royki ng		COLUMNS: Kj onn		
	1	2	ALL	
0	85.04	14.96	100.00	
	76.60	59.37	73.41	
	62.43	10.98	73.41	
1	71.74	28.26	100.00	
	23.40	40.62	26.59	
	19.08	7.51	26.59	
ALL	81.50	18.50	100.00	
	100.00	100.00	100.00	
	81.50	18.50	100.00	

CELL CONTENTS --
 % OF ROW
 % OF COL
 % OF TBL

Som synes røyker 23.4% av guttene, 40.6% av jentene⁵.

MISSing lager en egen kolonne/rad for "manglende verdi". Vanligvis er ikke "manglende verdi" med i tabuleringen — men hvis det er mange forekomster av "manglende verdi" kan det jo være interessant å se om disse adskiller seg fra resten av dataene — f.eks. for å se om det er noen fellestrekk ved folk som *ikke* har besvart et spørsmål på et spørreskjema.

NOALL fjerner kolonnen/raden med "ALL" fra tabellen. Grei å bruke hvis du ikke er interessert i disse opplysningene.

FREQuencies brukes hvis tallene i kolonnen ikke stemmer med antallet observerte. I stedet har du en ekstra kolonne som inneholder antall av hver observasjon (også kalt "vekten" av observasjonene). Et eksempel:

Kolonne C1 inneholder kodene for kjønn, dvs. 145 1-tall og 32 2-tall. Dette kunne i stedet vært lagt i to kolonner C11 og C12 slik:

C11	C12
1	145
2	32

Hvis du så gir kommandoen

TABLE C11;

⁵Årsaksforholdet mellom røyking og en rekke andre ting diskuteres heftig. Jeg vil i alle fall påstå røyking er årsak til en masse statistikk

FREQuenci es C12.

vil du få samme resultat som den første TABLE-kommandoen i dette kapittelet.

LAYOut *r k*

brukes for å forandre på standard-formatet for tabeller. Hvis du tabulerer mer enn to variable, kan du spesifisere hvor mange av variablene som skal være rader (*r*) og kolonner (*k*) i tabellene. I eksemplet nedenfor spesifiseres det at variabelen "Alder" skal være radene i tabellen, og at variablene "Kjønn" og "Røyking" skal være kolonnene. "Røyking", som er spesifisert sist, blir vist innenfor kjønn.

```
MTB > TABLE 'al der' 'kj onn' 'royki ng' ;
SUBC> LAYOut 1 2.
```

	1		2		ALL
	-----		-----		-----
	0	1	0	1	ALL
19	4	0	1	0	5
20	22	6	8	6	42
21	43	8	6	2	59
22	24	12	0	3	39
23	6	3	2	1	12
24	6	2	0	1	9
25	2	2	1	0	5
26	0	0	1	0	1
36	1	0	0	0	1
ALL	108	33	19	13	173

CELL CONTENTS --
COUNT

Ut fra denne tabellen kan du se at det var 22 mannlige ikke-røykere som var 20 år gamle blant studentene, 8 kvinnelige 20-årige ikke-røykere.

CHISquare [*n*] Foretar en Chi²-test for cellene i tabellen. *n* er en kode for hvor mye data som skal skrives ut. Hvis *n* = 1, skrives antallet observerte for hver celle ut (dette er standard), hvis *n* = 2, skrives forventet antall ut også, og hvis *n* = 3, skrives også standardisert residual for hver celle ut.

Se forøvrig eget punkt om Chi²-test (punkt ?, side 67)

Sannsynlighetsfordelinger

I dette kapitlet vil jeg ta for meg sannsynlighetsfordelinger og tilfeldige tall. Minitab kan operere med en rekke sannsynlighetsfordelinger — og benytte dem dels til å beregne sannsynligheten for at noe skal forekomme, dels for å generere tilfeldige tall som følger den fordelingen man ønsker.

Disse sannsynlighetsfordelingene spesifiserer du som subkommandoer til kommandoene PDF, CDF, INVCDF og RANDOM. Forvirret? Slapp av, etter noen eksempler er det hele så meget klarere.

5.2 Hvilke fordelinger finnes?

La oss begynne med en liste over de (av studenter) mest brukte sannsynlighetsfordelingene:

BERNoulli x kan bare benyttes sammen med RANDOM-kommandoen. Den lager tilfeldige tall, enten 0'er eller 1-tall, hvor sannsynligheten for å få 1 er lik x . Den egner seg med andre ord utmerket til å simulere hendelser hvor utfallet er enten 1 (suksess) eller 0 (fiasko). F.eks. vil en fordeling BERNoulli⁶ 0.5 tilsvare myntkast (hvor det jo er samme sannsynlighet for å få 1 (som f.eks. kan være krone) eller 0 (i dette tilfelle mynt)).

BINOMial $x y$ beregner sannsynligheter for binominalfordelinger, hvor x er antall ganger man forsøker, og y er sannsynligheten for suksess. En binominalfordeling tilsvarer trekking fra en endelig populasjon *med* tilbakelegging.

CHISquare x gir en Chi^2 (Kji-kvadrat) fordeling med x frihetsgrader.

DISCcrete Cx Cygir deg muligheten til å bestemme sannsynlighetsfordelinger (dvs. for diskrete hendelser) selv. Dette gjøres ved at du før du gir denne kommandoen har lagret utfallene i kolonne Cx, og sannsynlighetene for de enkelte utfall i kolonne Cy. For eksempel: Du har en terning som i stedet for feltene med 5 og 6 prikker har felt med 1 prikk, og i stedet for feltet med 4 prikker har et med 2 prikker. Dette vil si at sannsynlighetsfordelingen for et kast med denne terningen er

DISCcrete C20 C21

hvor C20 og C21 ser slik ut:

C20	C21
1	0.5
2	0.3333333
3	0.1666667

⁶Jacques Bernoulli, sveitsisk matematiker (1654-1705). Utga *Ars Conjectandi*, som inneholdt den første beskrivelse av "de store talls lov", posthumt i 1713. Han var en tidlig statistiker, og en av grunnleggerne av høyere matematikk. Den første statistiker var sannsynligvis den italienske legen Geronimo Cardano (1501-1576), som var spilllegal (og dessuten et matematisk geni med professorat ved universitetet i Praha). Han skrev *Liber de Ludo Aleae* (Håndbok i terningspill) i 1526, hvor han bl.a. beskrev addisjon av sannsynligheter. Denne boken ble imidlertid ikke utgitt før nesten 150 år etter (Campbell, 1982; Boyer, 1991).

F x y	gir en Fischer-fordeling (F-fordeling) for x er telleren (df_1) og y er nevneren (df_2).
INTEgers x y	gir en uniform fordeling med heltall f.o.m. x t.o.m. y . Dette vil si at det er lik sannsynlighet for å få hvert av heltallene f.o.m. x t.o.m. y . Et terningkast med en "ærlig" terning vil være en hendelse med en sannsynlighetsfordeling lik INTEgers 1 6.
NORMal μ σ	gir en normalfordeling med forventning = μ og standardavvik = σ . Hvis du ikke setter noe μ eller σ , brukes henholdsvis 0 og 1.
POISSon μ	gir en Poisson-fordeling ⁷ med forventning μ . Denne verdien må være større enn 0 og ikke større enn 50.
T x	gir en T-fordeling med x frihetsgrader.
UNIFORM x y	gir en uniform fordeling f.o.m. x t.o.m. y . Dette vil si at det er like stor sannsynlighet for å få en hvilken som helst verdi mellom x og y . Hvis du ikke setter noe x eller y , brukes 0 for x og 1 for y . Legg merke til at denne fordelingen er <i>kontinuerlig</i> , i motsetning til INTEger-fordelingen som er <i>diskret</i> .

Det finnes endel andre sannsynlighetsfordelinger også; se [HELP PDF](#) for en liste.

5.3 Å finne sannsynligheter

For å finne sannsynligheter har du 2 kommandoer tilgjengelig i Minitab: PDF (for punktsannsynligheter) og CDF (for kumulative sannsynligheter).

La oss se litt mer detaljert på dette:

5.3.1 Punktsannsynligheter

PDF (probability density function) gir deg **punktsannsynligheter** (eller, hvis det er snakk om kontinuerlige fordelinger, sannsynlighetstetthet). Dette vil si sannsynligheter for at en hendelse skal ha et enkelt utfall (f.eks. at et fotballag skal vinne 3 av 10 kamper).

Den generelle formen for PDF-kommandoen er

⁷Oppkalt etter Siméon-Denis Poisson (1781-1840), professor i matematikk ved Ecole Polytechnique i Paris, var bl.a. professor for Norges mest kjente matematiker, Niels Henrik Abel (1802-1829). Poisson, som arbeidet innenfor mange felter av anvendt matematikk og fysikk, publiserte denne distribusjonen, også kalt Poisson's lov om store tall, i *Recherches sur la probabilité des jugements* (1837). Denne går ut på at når n går mot uendelig, og p samtidig går mot 0, slik at produktet np er konstant, vil en binominal distribusjon nærme seg Poisson-distribusjonen. Siden Poisson-distribusjonen kun har et parameter, $\mu = np$, og dessuten er tilnærmet lik binominalfordelingen ved store datamengder, er den meget brukt til å beregne sannsynligheter for sjeldne, diskrete hendelser.

PDF [*kildekolonne*] [*målkolonne*]

"Kildekolonnen" inneholder observasjoner det skal beregnes sannsynligheter for. I stedet for kildekolonne kan du godt sette en konstant. Merk dette med PDF-kommandoen:

- målkolonne kan du benytte hvis du vil lagre de beregnede sannsynlighetene — hvis du ikke oppgir målkolonne blir resultatet skrevet ut på skjermen.
- Hvis du ikke oppgir noen spesiell sannsynlighetsfordeling (ved subkommando) bruker PDF en normalfordeling med $\mu = 0$ og $\sigma = 1$.
- Hvis du benytter fordelingene BINominal eller POISSon, trenger du ikke oppgi noen kildekolonne — resultatene kommer rett opp på skjermen i stigende rekkefølge.

Et eksempel på bruk av PDF:

Noe er binomisk fordelt med sannsynlighet for "treff" (p) = 0.3. Hva er så sannsynligheten for å få nøyaktig 8 "treff" hvis vi forsøker 20 ganger? For å finne punktsannsynligheter bruker vi PDF-kommandoen:

```
MTB > PDF 8;  
SUBC> BINominal 20 .3.
```

K	$P(X = K)$
8	0.1144

Hvis vi er interessert i sannsynlighetene for *alle* utfallene, utelater vi å spesifisere en enkeltverdi:

```
MTB > PDF;  
SUBC> BINominal 20 .3.
```

```
BINOMIAL WITH N = 20 P = 0.300000  
K          P(X = K)  
0          0.0008  
1          0.0068  
2          0.0278  
3          0.0716  
4          0.1304  
5          0.1789  
6          0.1916  
7          0.1643  
8          0.1144  
9          0.0654  
10         0.0308  
11         0.0120  
12         0.0039  
13         0.0010  
14         0.0002
```

Vi ser at ved 15 er sannsynligheten blitt så liten at den ikke kommer med i et tall med 4 desimaler — derfor markerer Minitab den med 0 og slutter å skrive ut tabellen (sannsynligheten er naturligvis ikke 0 — men den er så liten at den "praktisk talt" er det).

5.3.2 Kumulative sannsynligheter

CDF (cumulative distribution function) gir deg *kumulative sannsynligheter*. Dette kan f.eks. være at det samme fotballaget skal vinne 3 eller færre kamper av 10 mulige. . . Kommandoen gis på samme måte som PDF (se forrige avsnitt) — og de samme punkter gjelder for den.

Først et lite eksempel som viser CDF brukt uten at du på forhånd har en kolonne med tall det skal beregnes sannsynligheter for (m.a.o. fungerer kommandoen bare som oppslag i en fast tabell).

Hva er sannsynligheten for å få 8 eller færre "treff" fra den samme populasjonen (se PDF-eksemplet)? Dette finner vi slik:

```
MTB > CDF;
SUBC> BINOMIAL 20 .3.
```

```

BINOMIAL WITH N = 20 P = 0.300000
K  P( X LESS OR = K)
0  0.0008
1  0.0076
2  0.0355
3  0.1071
4  0.2375
5  0.4164
6  0.6080
7  0.7723
8  0.8867
9  0.9520
10 0.9829
11 0.9949
12 0.9987
13 0.9997
14 1.0000
```

Her ser vi at sannsynligheten for å få 8 eller færre "treff" er 0.8867. Vi ser også at Minitab slutter å skrive ut listen ved 14 "treff" — fordi sannsynligheten er så stor at den er 1 hvis vi runder av til 4 desimaler (og, som sagt før — den er naturligvis ikke 1 — bare nesten).

Så et eksempel hvor vi benytter CDF for å beregne de kumulative sannsynligheter for en rekke observasjoner vi har i en kolonne:

La oss si at vi har en kolonne C1 med disse tallene:

110 120 125 130 135 140 160 200

og vi betrakter dem som et mulig utvalg fra en normalfordelt populasjon med μ (populasjonsgjennomsnitt) = 130 og et standardavvik (σ) på 35. Hva er sannsynligheten for at vi skal få en verdi som er lik eller mindre for de enkelte verdiene i denne kolonnen? Eller, med andre ord, hvor stor del av arealet under normalfordelingskurven ligger til venstre for de oppgitte verdiene?

```
MTB > CDF c1;
SUBC> NORMal 130 25.
110. 0000    0. 2119
120. 0000    0. 3446
125. 0000    0. 4207
130. 0000    0. 5000
135. 0000    0. 5793
140. 0000    0. 6554
160. 0000    0. 8849
200. 0000    0. 9974
```

Som vi ser, er sannsynligheten for at en tilfeldig uttrukket observasjon fra denne populasjonen er 140 eller mindre lik 0.6554.

5.4 Å finne kritiske verdier

Som regel finner du kritiske verdier (dvs. verdier som utgjør grenser for hvorvidt du skal forkaste hypoteser eller ikke) ved å slå opp i en tabell. Minitab lar deg gjøre dette ved hjelp av INVCdf-kommandoen.

La oss si at du skal finne den kritiske verdien for Z i den standardiserte normalfordelingen ($\mu=0$, $\sigma=1$, $\alpha=0.05$). Denne fordelingen er *default*, dvs. den fordelingen Minitab benytter hvis du ikke oppgir noe annet:

```
MTB > INVCdf .95;
0. 9500    1. 6449
```

Vær oppmerksom på at INVCdf tar utgangspunkt i **ensidige** verdier — det tallet du oppgir er altså verdien **til venstre** for Z-verdien (dvs. $1 - \alpha$). Hvis du skal finne den tosidige verdien må du halvere signifikansnivået, f.eks. slik

```
MTB > INVCdf .975;
0. 9500    1. 9600
```

Nok et eksempel: La oss si at du skal finne den kritiske verdien for T med signifikansnivå 5% og 6 frihetsgrader. Du gir da kommandoen

```
MTB > INVCdf .95;
SUBC> T 6.
0. 9500    1. 9432
```

og får svaret 1.9432, som er den kritiske T-verdien.

Du kan, som med PDF- og CDF-kommandoene, i stedet for et enkelt tall oppgi en kolonne, og du kan oppgi en kolonne å lagre verdiene i.

5.5 Tilfeldige tall og utvalg

Innenfor tilfeldige tall og utvalg er følgende kommandoer aktuelle i Minitab:

- **RANDOM**-kommandoen, som produserer tilfeldige tall etter en spesifisert sannsynlighetsfordeling
- **BASE** — som benyttes når du vil produsere de samme tilfeldige tall om igjen
- **SAMPlE** — som benyttes når du vil trekke ut tilfeldige respondenter fra et materiale — m.a.o. trekke ut tilfeldige rader fra kolonne(r).

5.5.1 RANDOM-kommandoen

RANdOm-kommandoen genererer (lager) tilfeldige tall etter de fordelinger du spesifiserer. Et forklarende eksempel til å begynne med:

```
MTB > RANDom 1000 C1;  
SUBC> INTEger 1 6.
```

Med denne enkle kommandosekvensen får du Minitab til å foreta 1000 terningkast, og lagre utfallene av terningkastene (som altså blir heltall f.o.m. 1 t.o.m. 6) i kolonne C1. (Du kan jo ta tiden datamaskinen brukte på dette — og forsøke å forestille deg hvor lang tid det ville ta med en vanlig terning. . . .).

En rask tabulering bekrefter resultatet:

```
MTB > TABLe C1  
  
ROWS: C1  
  
      COUNT  
1      167  
2      139  
3      185  
4      175  
5      164  
6      170  
ALL    1000
```

Som du ser av utskriften, er det endel variasjon i antall ganger "terningen" har kommet opp med de ulike resultatene.

Hvis du ikke spesifiserer noen fordeling til RANDom-kommandoen (med en subkommando), bruker Minitab en normalfordeling med $\mu = 0$ og $\sigma = 1$.

Måten Minitab (i likhet med en rekke andre datapakker) genererer tilfeldige tall på, er at den har liggende en lang "streng" av tall, som er trukket ut gjennom en tilfeldig prosess (denne lange tallstrengen igjen simulert ved en matematisk funksjon). Når du gir RANDom-kommandoen, starter Minitab på et tilfeldig sted i denne tallrekken, leser tallene, og bearbeider dem slik at de får den fordelingen du har spesifisert. Utgangspunktet for denne metoden er at det er svært vanskelig å generere tall som er tilfeldige ved hjelp av en enkel funksjon — derfor bruker man "lagrede" tall.

Den observante leser vil nå stille spørsmålet: hvordan finner så Minitab det "tilfeldige" startstedet i denne tallrekken? Metoden for dette varierer litt mellom de ulike datamaskintypene, men utgangspunktet er at Minitab leser av klokkeslettet fra maskinens interne klokke, og deretter benytter f.eks. antall tusendels sekunder som "startsted" i tallrekken. Det skal godt gjøres å gi Random-kommandoen på det samme tusendels sekund to ganger etter hverandre.

Hva så hvis du ønsker å generere den samme rekken av tilfeldige tall to ganger etter hverandre? Det kan du gjøre ved hjelp av **BASE**-kommandoen, som gis slik:

`BASE x`

hvor x er et hvilket som helst tall. Dette tallet vil da bli konvertert til en startposisjon i tallrekken. BASE virker helt til du går ut av Minitab — eller til du gir en ny BASE-kommando.

5.5.2 Et eksempel: De Store Talls Lov.

I det følgende eksemplet skal vi se på "de store talls lov", også kalt sentralgrenseteoremet, som sier at

når antallet utvalg øker, vil fordelingen av gjennomsnittet av utfallene i hvert utvalg nærme seg normalfordelingen

eller, for å si det på en annen måte:

Når verdiene til x i en stikkprøvepopulasjon ikke er normalfordelte, vil likevel fordelingen til stikkprøvegjennomsnittet x være det (Wenstøp, 1987).

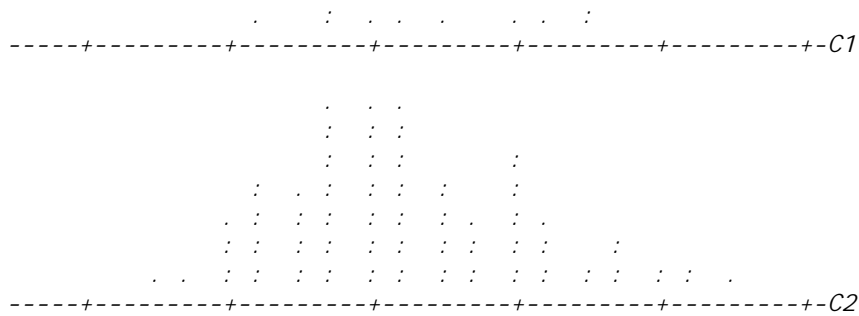
La oss ta en liten kikk på hvordan dette fungerer ved hjelp av RANDom-kommandoen. Vi velger oss en enkel fordeling: Binominal med 50% sjanse for å få enten 1 eller 0. Vi tar utvalg på 50, og gjør dette 10, 100 og 1000 ganger:

```
MTB > RANDom 10 c1;
SUBC> BI Nomi nal 50 .5.
MTB > RANDom 100 c2;
SUBC> BI Nomi nal 50 .5.
MTB > RANDom 1000 c3;
SUBC> BI Nomi nal 50 .5.
```

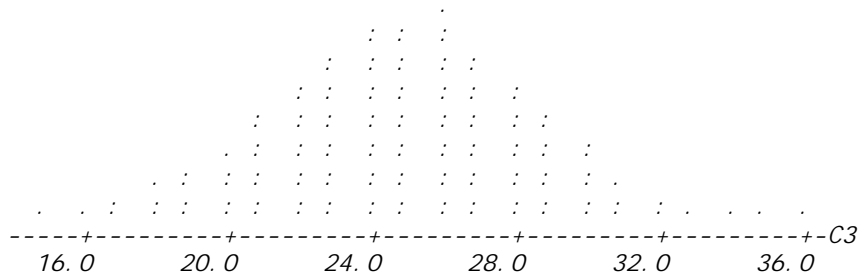
Deretter bruker vi DOTPlot-kommandoen (se side 76) for å vise resultatet:

```
MTB > DOTPl ot c1-c3;
```

SUBC> SAME.



Each dot represents 8 points



Som synes fordeler stikkprøvegjenomsnittene seg mer og mer som en normalfordeling etterhvert som antallet "trekk" øker. Legg merke til at det hele tiden er snakk om *gjennomsnittet* av hvert av utvalgene, dvs. gjennomsnittet av 50 trekk.

5.5.3 SAMPLE-kommandoen

Hvis du har en kolonne med data, og ønsker å trekke ut et tilfeldig utvalg fra denne kolonnen, kan du gjøre dette med kommandoen

*SAMPl*e *antall kildekolonne(r) målkolonne(r)*

eller f.eks.

*SAMPl*e 10 C3 C13

som betyr "trekk ut 10 av dataene i C3, legg dem inn i C13". Uttrekningene blir foretatt **uten** tilbakelegging — dvs. du kan ikke trekke den samme respondenten (raden) flere ganger. Hvis du ønsker at trekningene skal foregå *med* tilbakelegging, må du bruke subkommandoen REPLace.

Du kan benytte flere kilde- og målkolonner på samme måte som COPY-kommandoen — f.eks. til å trekke tilfeldige personer fra flere som du har stilt spørsmål til. Dette er en teknikk som gjerne benyttes når man skal stille "ladede" spørsmål — eller du kan jo benytte den til å trekke ut et lite antall data for å "prøvekjøre" analyser på.

5.5.4 Et simuleringseksempel: enarmet banditt

RANDOM- og SAMPLE-kommandoene benyttes ofte ved simuleringer. La oss se litt på et eksempel hvor vi simulerer en *enarmet banditt*.

En enarmet banditt er en spilleautomat som de fleste sikkert har sett i hvert fall et eksemplar av, om ikke annet så på bilder fra Las Vegas (eller danskebåten). Den ser ut som en kasse med tre små vinduer plassert ved siden av hverandre. I disse vinduene er det tre hjul, plassert slik at man gjennom vinduene ser ytterkanten av hjulet (slik man ville sett slitebanen på tre bildekk plassert stående ved siden av hverandre). På hjulene er det trykket symboler. Antallet varierer, men det er de samme symbolene på hvert hjul. Ved å trekke ned en arm (etter at man har puttet på penger naturligvis — maskinen heter ikke "banditt" for ingenting) settes hjulene i bevegelse. Ved hjelp av en fjærmekanisme stanses hjulene — og man får premier ut fra hvilke symboler som står ved siden av hverandre.

La oss tenke oss en slik enarmet banditt med symboler — og at vi skal undersøke muligheten for gevinst. Vi tenker oss at det er 8 forskjellige symboler på hvert hjul — et eple⁸ og 7 andre symboler. Premieringen er som følger:

Vindu 1	Vindu 2	Vindu 3	Utbetaling
Eple	Eple	Eple	250.-
Eple	Eple	Ikke eple	25.-
Eple	Ikke eple	Ikke eple	1.-

Hva er forventet gevinst av 1000 spill á kr 1.-?

Dette kan regnes ut manuelt, men vi forsøker oss med en simulering:

Først lager vi to kolonner som inneholder sannsynligheten for eple/ikke eple:

```
MTB > READ C10 C11
DATA> 1 .125      (sannsynligheten for "eple" = 1/8 = .125)
DATA> 0 .875      (sannsynligheten for "ikke eple")
DATA> END
MTB > _
```

Deretter gjennomfører vi simuleringen:

```
MTB > RANDom 1000 C1-C3;      (1000 trekk i "armen". . . . .)
SUBC> DIScrete C10 C11.      (. . . med sannsynligheter.)
```

Så beregner vi resultatet:

```
MTB > LET C4 = (C1 EQ 1) AND (C2 EQ 1) AND (C3 EQ 1)
MTB > LET C5 = (C1 EQ 1) AND (C2 EQ 1) AND (C3 NE 1)
MTB > LET C6 = (C1 EQ 1) AND (C2 NE 1) AND (C3 EQ 1)
```

Disse linjene lager tre nye kolonner: i C4 vil det være nuller, bortsett fra i de tilfelle hvor det er "eple" i alle vinduene, i C5 vil det være nuller bortsett fra når det er et eple i hvert av de to første vinduene osv. Det som står igjen nå, er å beregne resultatet:

⁸Fordi disse symbolene ofte er forskjellige slags frukt, kalles enarmede banditter ofte *fruit machines* på engelsk.

```

MTB > LET K4 = SUM(C4)
MTB > LET K5 = SUM(C5)
MTB > LET K6 = SUM(C6)
MTB > LET K1 = K4 * 250 + K5 * 25 + K6 - 1000

```

og å skrive det ut:

```

MTB > PRINT K4 K5 K6 K1
K4          3.00000
K5          14.0000
K6          8.00000
K1          108.000

```

Eller, med andre ord, vi vant kr 108,- på å spille på denne banditten. Men vil dette gjelde i det lange løp?

Det er to måter vi kan finne ut det på. Vi kan beregne forventet gevinst nøyaktig:

$$\begin{aligned}
& [P(3Epler) * 250] + [P(Eple / Ikkeple / Eple) * 25] + P(Eple / Ikkeple / Ikkeple) = \\
& \left[\left(\frac{1}{8}\right)^3 * 250 \right] + \left[\left(\frac{1}{8}\right)^2 * \left(\frac{7}{8}\right) * 25 \right] + \left[\frac{1}{8} * \left(\frac{7}{8}\right) \right] = \\
& [0.001953125 * 250] + [0.013671875 * 25] + 0.095703125 \approx \underline{0.9258}
\end{aligned}$$

Som synes vil man i det lange løp tape ca. 7.4% av innsatsen.

Vi kan også finne ut dette ved å gjøre simuleringen en rekke ganger. Et eksempel på hvordan du enkelt kan kjøre mange simuleringer er vist i kapittelet om programmering av Minitab (kapittel ?, side 105ff.).

Du vil her kunne se at det av og til er mulig å vinne endel, trass i at forventet utbetaling er mindre enn innsatsen. Det er denne muligheten som holder liv i en hel del casinoer rundt omkring i verden.

Tester og konfidensintervall

Det følgende er en kort forklaring på noen av Minitabs kommandoer for hypotesetester og konfidensintervall. Jeg har ikke detaljert forklart grunnlaget for de ulike testene — hvis du vil ha en nærmere forklaring, se en nesten hvilken som helst lærebok i statistikk. Felles for alle disse testene er at de forutsetter at populasjonen er normalfordelt, eller i alle fall tilnærmet normalfordelt.

5.6 Z-test og -intervall

Disse testene benyttes på utvalg hvor populasjonens standardavvik (σ , sigma) er kjent. I praksis er dette meget sjelden tilfelle, men det er ikke helt ukjent i statistikkoppgaver. . . .

La oss først se på konfidensintervallene. Vi tar utgangspunkt i en kolonne som inneholder endel personer vekt (her: C3), og betrakter den som trukket ut fra en populasjon hvor σ (dvs. *populasjonens* standardavvik) er lik 9.5 kg. For å få et 95 % konfidensintervall for populasjonens gjennomsnitt (μ) gir vi kommandoen

```
MTB > ZINTErval 9.5 C3
```

```
THE ASSUMED SIGMA =9.50
```

	N	MEAN	STDEV	SE MEAN	95.0 PERCENT C. I.
Vekt	212	69.413	9.464	0.652	(68.132, 70.693)

Med andre ord: vi kan med 95% sikkerhet si at populasjonens gjennomsnitt ligger mellom 68,1 og 70,7 kg (under antagelse om et standardavvik på 9.5 kg).

Den generelle formen på kommandoen er

```
ZINTErval [k]  $\sigma$  kolonne(r)
```

hvor k er konfidensnivået, og σ er populasjonens (antatte) standardavvik. Konfidensnivået (som er 95% hvis du ikke spesifiserer noe annet) kan skrives som % (f.eks. 80) eller en desimal (f.eks. 0.80). Minitab regner i begge tilfelle med at det er snakk om konfidensintervaller på mellom 1% og 99%.

En *Z-test* tester hypoteser om en populasjons gjennomsnitt μ_0 (my) gitt at populasjonens standardavvik σ (sigma) er kjent.

Den generelle formen for ZTEST-kommandoen i Minitab er

```
ZTEST  $\mu_0$   $\sigma$  kolonne(r)
```

hvor kolonnen(e) inneholder de utvalg du ønsker å teste.

La oss ta et enkelt eksempel først (vi bruker fortsatt kolonne C3 som inneholder vekten til endel personer):

```
MTB > ZTEST 72 9.5 C3
```

TEST OF MU = 72.000 VS MU N. E. 72.000
 THE ASSUMED SIGMA = 9.50

	N	MEAN	STDEV	SE MEAN	Z	P VALUE
Vekt	212	69.413	9.464	0.652	-3.97	0.0001

Her tester vi en hypotese om at populasjonens gjennomsnittsvikt er 72 kg. Vi antar en σ (populasjonens standardavvik) på 9.5 kg. Minitab returnerer en Z-verdi på -3.97 . Kritisk verdi på 5% signifikansnivå (tosidig test) er 1.96^9 , og H_0 kan dermed forkastes. Alternativt kan du si at H_0 forkastes fordi p-verdien (P VALUE) er mindre enn α (signifikansnivået).

For å gjøre *ensidige* Z-tester benytter vi subkommandoen ALTErnative. Denne kommandoen kan ha to former:

ALTErnative = 1 tester $H_0: \mu > \mu_0$

ALTErnative = -1 tester $H_0: \mu < \mu_0$

Denne subkommandoen benyttes på en rekke av hypotesetestingskommandoene i Minitab.

Et eksempel: Vi ønsker å teste en hypotese H_0 at μ (populasjonens gjennomsnitt) er 70 Kg mot en hypotese H_1 om at μ er mindre enn 70 kg. Vi antar at σ (populasjonens standardavvik) er 9.5 kg. Signifikansnivå setter vi til 5%. Vi får denne sekvensen:

```
MTB > ZTEST 70 9.5 C3;
SUBC> ALTErnative = -1.
```

TEST OF MU = 70.000 VS MU L. T. 70.000
 THE ASSUMED SIGMA = 9.50

	N	MEAN	STDEV	SE MEAN	Z	P VALUE
Vekt	212	69.413	9.464	0.652	-0.90	0.18

Som vi ser, blir den observerte Z-verdien -0.90 . Den kritiske Z-verdien finner vi ved hjelp av INVCdf-kommandoen slik:

```
MTB > INVCdf .95;
0.9500 1.6449
```

Siden observert verdi (-0.90) ligger innenfor kritisk verdi (-1.6449) må H_0 beholdes — vi kan ikke med 95% sikkerhet si at populasjonens gjennomsnittsvikt er under 70 kg. (Hvis du tegner et normalfordelingsdiagram her så blir det hele så meget klarere. . . .).

5.7 T-test og -intervall

⁹En liten utfordring: finn denne ved hjelp av INVCdf-kommandoen!

T-tester ligner på Z-tester — med unntak av at de ikke forutsetter at σ (populasjonens standardavvik) er kjent. Dette gjør at det vanligvis er T-tester og T-konfidensintervaller du vil benytte. T-intervaller vil alltid være videre enn Z-intervaller, og T-tester mer tilbøyelige til å forkaste H_0 enn Z-tester, fordi μ må estimeres.

Den generelle formen for et T-konfidensintervall er i Minitab

`TINTerval [k] kolonne(r)`

hvor k er konfidensnivået (95% er default¹⁰).

La oss ta et eksempel (stadig vekk benytter vi de samme data som under Z-testene). Vi vil ha et 95% konfidensintervall for populasjonens gjennomsnitt, ut fra et antall observasjoner i C3 (kroppsvekt):

`MTB > TINTerval C3`

	<i>N</i>	<i>MEAN</i>	<i>STDEV</i>	<i>SE MEAN</i>	<i>95.0 PERCENT C. I.</i>
<i>Vekt</i>	212	69.413	9.464	0.650	(68.131, 70.694)

Utskriften skulle tale for seg selv — men legg merke til at konfidensintervallet er videre enn Z-intervallet. Forskjellen er imidlertid liten — jo flere observasjoner, jo mindre forskjell mellom Z- og T-testen. Vi har 212 observasjoner, som er ganske mange i denne sammenheng.

En *T-test* tester hypoteser på samme måte som en Z-test; men som sagt, du trenger ikke kjenne σ (populasjonens standardavvik) på forhånd.

Den generelle formen for TTEST-kommandoen er

`TTEST [m] kolonne(r)`

hvor m er den μ (populasjonsgjennomsnitt) du ønsker å teste. Hvis du ikke skriver noen m , bruker Minitab 0.

Vi tar for oss det samme eksemplet vi benyttet i Z-testen: Vi skal teste (på 5% nivået) en hypotese H_0 om at gjennomsnittsvekten for populasjonen er 70 kg mot en hypotese H_1 om at den er **mindre enn** 70 kg. Vi benytter da subkommandoen `ALTErnative = -1` (se under `ZTEST`-kommandoen).

`MTB > TTEST 70 c3;`
`SUBC> ALTErnative = -1.`

`TEST OF MU = 70.000 VS MU L. T. 70.000`

	<i>N</i>	<i>MEAN</i>	<i>STDEV</i>	<i>SE MEAN</i>	<i>T</i>	<i>P VALUE</i>
<i>Vekt</i>	212	69.413	9.464	0.650	-0.90	0.18

¹⁰Default er et datauttrykk som betyr omtrent det samme som *standard*; dvs. den verdi som benyttes hvis du ikke oppgir noe annet.

Som vi ser er P VALUE 0.18, altså større enn signifikansnivået. Dermed må vi beholde H_0 . Vi ser også at oppnådd T-verdi er -0.90 . Kritisk verdi (med $(N-1)$, dvs. 211 frihetsgrader) finner vi med `INVCdf`-kommandoen, slik:

```
MTB > INVCdf .95;
SUBC> T 211.
      0.9500      1.6521
```

Siden observert T-verdi (-0.9) er større enn kritisk verdi (-1.6521) må H_0 beholdes — vi kan ikke med 95% sikkerhet si at populasjonens gjennomsnittsvekt er under 70 kg.

5.8 Test av samsvar mellom to utvalg (inferens)

`TWOSample`-kommandoen tester hypoteser om hvorvidt to populasjoner har samme gjennomsnitt (μ), basert på to utvalg.

Den generelle formen er

```
TWOSample [k] kolonne1 kolonne2
```

hvor k er konfidensnivået (standard er 95%). `TWOSample` har to subkommandoer: `ALTErnative` og `POOLed`. `ALTErnative` spesifiserer H_1 , på denne måten:

```
uten ALTErnative      :       $H_1: \mu_1 \neq \mu_2$  (dvs. tosidig test)
ALTErnative = 1      :       $H_1: \mu_1 > \mu_2$ 
ALTErnative = -1     :       $H_1: \mu_1 < \mu_2$ 
```

Hvis du gir subkommandoen `POOLed`, antar du at populasjonene har samme varians (σ^2).

Eksempel på bruk av `TWOSample`:

Vi skiller dataene for kroppshøyde for menn og kvinner i to separate kolonner, kaller dem 'Cm-Menn' og 'Cm-Kvinn', og tester om vi på grunnlag av disse utvalgene kan si at $\mu(\text{høyde menn}) = \mu(\text{høyde kvinner})$. Vi foretar testen ensidig, med $H_1: \mu(\text{menn}) > \mu(\text{kvinner})$. Først fordeler vi dataene i to kolonner etter kjønn, og navngir kolonnene

```
MTB > UNSTack 'Hoyde' C11 C12;
SUBC> SUBScri pts 'Kjonn' .
MTB > NAME C11 'Cm-Menn' C12 'Cm-Kvi nn'
```

Deretter foretar vi testen, først uten å anta at variansen er lik i de to populasjonene:

```
MTB > TWOSampl e C11 C12;
SUBC> ALTErnative = 1.
```

```
TWOSAMPLE T FOR Cm-Menn VS Cm-Kvi nn
          N      MEAN      STDEV      SE MEAN
Kg-Menn  145    182.38      6.47      0.54
Kg-Kvi nn  32    170.09      4.95      0.88
```

95 PCT CI FOR MU Cm-Menn - MU Cm-Kvinn: (10.23, 14.34)

TTEST MU Cm-Menn = MU Cm-Kvinn (VS GT):
T= 11.96 P=0.0000 DF= 57

Deretter prøver vi med POOLED varians:

```
MTB > TWOSample C11 C12;  
SUBC> ALTErnati ve = 1;  
SUBC> POOLed.
```

```
TWOSAMPLE T FOR Cm-Menn VS Cm-Kvinn  
N MEAN STDEV SE MEAN  
Kg-Menn 145 182.38 6.47 0.54  
Kg-Kvinn 32 170.09 4.95 0.88
```

95 PCT CI FOR MU Cm-Menn - MU Cm-Kvinn: (9.88, 14.69)

TTEST MU Cm-Menn = MU Cm-Kvinn (VS GT):
T= 10.10 P=0.0000 DF= 175

POOLED STDEV = 6.23

I begge tilfelle forkastes H_0 på 5%-nivået (kritisk T-verdi er hhv. 1.67 og ca 1.65): gjennomsnittlig høyde er signifikant større for menn enn for kvinner.

En variant av TWOSample-kommandoen er TWOT. Denne kommandoen er akkurat lik TWOSample, bortsett fra at begge stikkprøvene kan befines seg, usortert, i samme kolonne. Den generelle formen er

TWOT [k] datakolonne kategorikolonne

hvor k er konfidensnivå (95% standard), "datakolonne" er kolonnen med dataene, og "kategorikolonne" er en kolonne som inneholder koder for hvilke grupper dataene fordeler seg på (f.eks. kjønn). En tilsvarende test som den i eksempelet med TWOSample ovenfor ville se slik ut:

```
MTB > TWOT C2 C1
```

```
TWOSAMPLE T FOR Hoyde  
Kjonn N MEAN STDEV SE MEAN  
2 32 170.09 4.95 0.88  
1 145 182.38 6.47 0.54
```

95 PCT CI FOR MU 2 - MU 1: (-14.34, -10.23)

TTEST MU 2 = MU 1 (VS NE): T= -11.96 P=0.0000 DF= 57

Også TWOT-kommandoen kan benytte subkommandoene ALTErnati ve og POOLed.

5.9 Kji-kvadrat-analyser

Kji-kvadrat (Chisquare, Chi^2) tester er en analyseform som har en rekke anvendelsesområder — bl.a. å finne signifikante forskjeller i andeler. I Minitab kan du foreta Chi^2 -tester på to måter: som en egen kommando, og som en subkommando til TABLE. Hvem av disse du benytter, er avhengig av hva slags data du har. Vi begynner med et eksempel med CHISquare som subkommando til TABLE:

Er det noen signifikant forskjell mellom andelen jenter som røyker og andelen gutter? Vi benytter studentdataene fra TABLE-kapittelet og setter opp disse hypotesene (signifikansnivå 5%):

- H_0 : Relativt like mange gutter som jenter røyker.
 H_1 : Det er forskjell mellom andelene jenter og gutter som røyker.

Deretter gir vi disse Minitab-kommandoene:

```
MTB > TABLE c7 c1;
SUBC> CHI Square 3;
SUBC> NOALI 11.

      ROWS: Royki ng      COLUMNS: Kj onn
           1           2
0      108           19
      103.51      23.49
       0.44      -0.93

1       33           13
      37.49      8.51
      -0.73      1.54

CHI -SQUARE =      3.962  WITH D. F. =      1

CELL CONTENTS --
                COUNT
                EXP FREQ
                STD RES
```

Som vi ser, er Chi^2 -tallet 3.962. Kritisk verdi finner vi ved INVCdf-kommandoen (forutsetter signifikansnivå 5%):

```
MTB > INVCdf .95;
SUBC> CHI Square 1.
      0.9500      3.8415
```

Som vi ser er kritisk verdi mindre enn observert verdi, og H_0 forkastes. Det er med andre ord (såvidt) signifikant forskjell i andelen røykere blant gutter og jenter.

¹¹NOALI-subkommandoen (som fjerner ALL-kolonnen og -raden fra tabellen) er brukt fordi vi ikke trenger å vite totalene her. Det er alltid "godt håndverk" ikke å ha med mer informasjon enn strengt nødvendig.

Som vi ser, er CHISquare-subkommandoen (under TABLE) gitt med et parameter: 3. Dette korresponderer med de dataene som blir skrevet ut i hver celle i tabellen. Hvis du gir subkommandoen med 1 (eller ingenting, siden 1 er standard), skrives kun antall innen hver celle ut. Hvis du skriver CHISquare 2, får du med et tall til i hver celle, nemlig *forventet* antall. Og hvis du gjør som i eksemplet overfor får du i tillegg med standardiserte residualer for hver celle. Standardiserte residualer

$$\text{Standardisertresidual} = \frac{\text{observertantall} - \text{forventetantall}}{\sqrt{\text{forventetantall}}}$$

beregnes etter denne formelen:

CHISquare finnes også som egen kommando. Denne kommandoen krever at du kjenner antallet i de enkelte andelene først — men ofte er det jo nettopp det du gjør. La oss si at vi hadde foretatt tabuleringen først (gutter: 108 røykere, 33 ikke-røykere; jenter: 19 – 13). Vi leser først inn dataene i to (korte) kolonner:

```
MTB > READ C11 C12
DATA> 108 19
DATA> 33 13
DATA> END
```

2 ROWS READ

Deretter gir vi CHISquare-kommandoen:

```
MTB > CHI Square C11 C12
```

Expected counts are printed below observed counts

	C11	C12	Total
1	108	19	127
	103.51	23.49	
2	33	13	46
	37.49	8.51	
Total	141	32	173

Chi Sq = 0.195 + 0.859 +
0.538 + 2.371 = 3.962

df = 1

Som du ser, får vi stort sett den samme informasjon — bortsett fra at Chi²-tallet er brutt opp i sine enkelte komponenter — og at standardisert residual mangler.

5.10 Variansanalyser

Variansanalyse (analysis of variance, ofte referert til som ANOVA) tar sikte på å teste hypoteser om forskjeller mellom gjennomsnittene i to eller flere sett av observasjoner. Minitab har en meget kraftig kommando, ANOVA, som tillater avanserte analyser av varians. I denne manualen vil jeg imidlertid kun omtale variansanalyse basert på én variabel ("one-way" analysis of variance).

Jeg vil benytte et eksempel fra Berenson & Levine (1979), hvor vi ser på levetiden (målt i 1000 miles) for bildekk produsert ved tre forskjellige arbeidsskift:

C1 Dag	C2 Etterm.	C3 Natt
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

Vi tester følgende hypotese:

$$H_0: \mu_{\text{Dag}} = \mu_{\text{Etterm.}} = \mu_{\text{Natt}}$$

H_1 : Minst et gjennomsnitt er forskjellig fra de andre.

Vi bruker kommandoen AOVOneway (Analysis of Variance Oneway):

```
MTB > AOVOneway C1 - C3
```

ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS	F	p
FACTOR	2	47.164	23.582	25.60	0.000
ERROR	12	11.053	0.921		
TOTAL	14	58.217			

INDIVIDUAL 95 PCT CI'S FOR MEAN
BASED ON POOLED STDEV

LEVEL	N	MEAN	STDEV
Day	5	24.930	1.032
Evening	5	22.610	0.882
Night	5	20.590	0.959

POOLED STDEV = 0.960

Vi får en observert F-verdi på 25.60. Den kritiske F-verdien (på 1% signifikansnivå) finner vi med INVCdf-kommandoen:

```
MTB > invcdf .99;
SUBC> f 2 12.
0.9900 6.9266
```

Som vi ser, er observert F-verdi større enn kritisk verdi. Dette betyr at H_0 kan forkastes — det er reell forskjell i gjennomsnittene for de tre skiftene.

En annen variant av AOVOneway er ONEWayaoov, (dette begynner å bli en smule kryptisk), eller ONEWay. Det er den samme typen analyse, men dataene har et annet format. Hvis vi går ut fra eksemplet med bildekkene ovenfor: I stedet for å ha tre kolonner med dataene for hvert skift i hver, ville vi hatt en kolonne med dataene, og en kolonne med 1 (dag), 2 (ettermiddag) og 3 (natt), på denne måten:

C1	C2
Levetid	Skift

25.40	1
26.31	1
24.10	1
23.74	1
25.10	1
23.40	2
21.80	2
23.50	2
22.75	2
21.60	2
20.00	3
22.20	3
19.75	3
20.60	3
20.40	3

Vi kan da gi kommandoen

MTB > ONEWAY c1 c2

ANALYSIS OF VARIANCE ON Levetid

<i>SOURCE</i>	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>
<i>Ski ft</i>	2	47.164	23.582	25.60	0.000
<i>ERROR</i>	12	11.053	0.921		
<i>TOTAL</i>	14	58.217			

*INDIVIDUAL 95 PCT CI'S FOR MEAN
BASED ON POOLED STDEV*

<i>LEVEL</i>	<i>N</i>	<i>MEAN</i>	<i>STDEV</i>	
1	5	24.930	1.032	-----+-----+-----+-----+----- (-----*-----)
2	5	22.610	0.882	(-----*-----)
3	5	20.590	0.959	(-----*-----) -----+-----+-----+-----+-----
<i>POOLED STDEV =</i>		0.960		20.0 22.0 24.0 26.0

Grafikk

Minitab har i utgangspunktet begrensede grafikkmuligheter¹². Det benyttes såkalt semigrafikk, som er grafikk ved hjelp av vanlige bokstaver og tegn. Fordelen med dette er at du kan generere grafikk på alle slags skjermer og skrivere — noe som gjør at Minitab kan kjøres på mange maskintyper. Ulempen er at grafikken er nokså primitiv — noe som vil fremgå av eksemplene.

Imidlertid er grafikken i Minitab slett ikke bortkastet — den er utmerket i sammenheng med f.eks. regresjonsanalyse. Man benytter da grafikken for å få en oversikt over sammenhengene mellom to kolonner — ikke, som grafikk vanligvis benyttes, for å fremstille data på en "appetittelig" måte.

Jeg vil her beskrive kommandoene

PLOT	som gir todimensjonale punktdiagrammer.
HISTogram	som gir histogrammer.
DOTPlot	som gir en annen form for histogrammer.
BOXPlot	som gir en visualisering av begrepet konfidensintervaller.
STEM	som gir en mer presis histogramtype.

Jeg tar utgangspunkt i "standardutseendet" på grafene. Det er imidlertid anledning ved hjelp av endel subkommandoer å tilpasse størrelse, intervaller o.l. på grafene — men dette vil i praksis svært sjelden bli brukt. Hvis du er ute etter mer spesialiserte grafer, vil det som regel lønne seg å overføre dataene fra Minitab inn i et regneark eller grafikkprogram — for så å benytte seg av de adskillig mer avanserte grafikkmulighetene du finner der.

Fra og med versjon 7 kan du "fikse på" grafer med subkommandoene

- TITLE som setter inn tekst sentrert over en graf. Du kan ha inntil tre TITLES på en graf (de kommer da sentrert under hverandre).
- FOOTnote som setter inn tekst sentrert under en graf. Du kan ha inntil to FOOTnotes samtidig på en graf.
- XLABEL som setter tekst sentrert under X-aksen.
- YLABEL som setter tekst sentrert på Y-aksen (skrives ut loddrett).

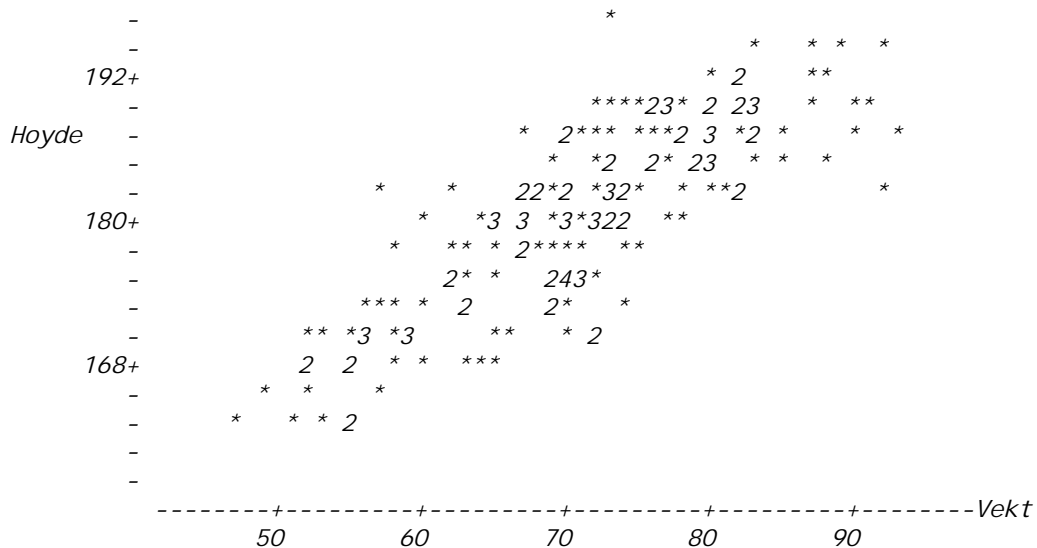
Den teksten du spesifiserer erstatter kolonnenavn eller -nummer.

5.11 PLOT (punktdiagrammer)

PLOT kan du benytte for å se om det er noe mønster (og derved noen sammenheng) mellom to variable. La oss ta et banalt eksempel, hvor vi plotter kroppshøyde (C2) mot kroppsvekt (C3).

```
MTB > PLOT C2 C3
```

¹²Det finnes grafikk av god kvalitet både for PC, Macintosh, IBM stormaskin og DEC VAX (som er de fire datamaskintypene som er behandlet i denne manualen). Imidlertid er disse grafikktypene såpass spesielle for hver datamaskintype (de avhenger blant annet av hva slags skjerm eller terminal du har) at jeg har funnet det mest riktig å ikke behandle dem detaljert i denne manualen. Det henvises til kapitlene om de enkelte maskintyper, kommandoene HELP OVERview 8 og HELP COMMands 7, samt manualer som følger med Minitab for de ulike maskintypene.



Som vi ser av denne grafen, er det en sterk samvariasjon mellom kroppshøyde og kroppsvekt — uten at det burde komme som en overraskelse på noen.

PLOT-kommandoen markerer "treffpunkter" på tre ulike måter:

- * markerer en enkeltobservasjon. Av diagrammet ovenfor ser vi at det f.eks. kun var en observasjon av en person med kroppshøyde 162 cm (eller rettere sagt, mellom 159.5 og 163.5 cm) og kroppsvekt 48 kg.
- 2–9 markerer punkter i diagrammet hvor det er mer enn én observasjon. Tallet angir da antallet observasjoner.
- + markerer punkter i diagrammet hvor det er mer enn 9 observasjoner (ingen punkter i diagrammet ovenfor).

Minitab bestemmer selv intervallene langs aksene (dvs. hvor "delestreke" skal gå). Du kan imidlertid selv spesifisere dette ved hjelp av subkommandoene XINCrement (intervall mellom delestreker langs X-aksen), XSTArt (startverdi X-aksen), YINCrement og YSTArt.

5.12 HISTogram

HISTogram-kommandoen lager (naturligvis) histogrammer. Et histogram er et søylediagram som gir en oversikt over fordelingen av observasjoner. Det er egnet for raskt å få en ide om fordelingen av et sett av data.

La oss ta et eksempel, hvor vi ser på kroppshøyde for et antall studenter (disse dataene ligger i kolonne 'Hoyde'):

```
MTB > HI STogram 'Hoyde'
```

```
Hi stogram of Hoyde    N = 177
```

<i>Mi dpoi nt</i>	<i>Count</i>	
164	5	*****
168	12	*****
172	25	*****
176	19	*****
180	29	*****
184	37	*****
188	28	*****
192	17	*****
196	5	*****

Minitab starter med å oppgi at det finnes 177 observasjoner og at hver "*" i diagrammet representerer 2 observasjoner. Deretter kommer histogrammet.

Hvis du ser litt nøyere på dette histogrammet ser du at det ikke er symmetrisk om midten. En grunn til dette kan være at det er både menn og kvinner med i observasjonene. Det kan derfor være lurt å se på menn og kvinners kroppshøyde hver for seg — for å se om kroppshøyden ligner mer på normalfordelingen da.

Dette kan du gjøre med subkommandoen BY. Denne subkommandoen gir deg et histogram for hver "kategoriseringsvariabel" (i dette tilfelle kjønn) som ligger i den kolonnen som BY viser til. La oss se dette i praksis:

```
MTB > HIStogram 'Hoyde' ;
SUBC> BY 'Kjonn' .
```

Hi stogram of Hoyde *Kjonn = 1* *N = 145*

<i>Mi dpoi nt</i>	<i>Count</i>	
164	0	
168	1	*
172	16	*****
176	16	*****
180	26	*****
184	36	*****
188	28	*****
192	17	*****
196	5	*****

Hi stogram of Hoyde *Kjonn = 2* *N = 32*

<i>Mi dpoi nt</i>	<i>Count</i>	
164	5	*****
168	11	*****
172	9	*****
176	3	***
180	3	***
184	1	*
188	0	
192	0	
196	0	

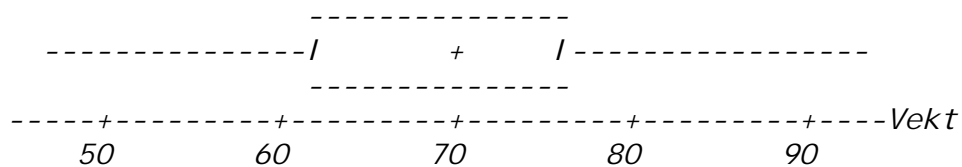
Som du ser, får vi to histogrammer (fordi dataene i kolonnen 'Kjonn', naturlig nok, bare var av to typer) — i dette tilfelle er det øverste for menn og det nederste for kvinner. Her ser vi at det *er* forskjeller mellom de to grupperingene — samt at de hver for seg får en fordeling som ligger nærmere den ideelle normalfordelingskurven enn den de hadde tilsammen.

5.13 BOXPlot

BOXPlot gir på en enkel og rask måte et visuelt bilde av innholdet i en eller flere kolonner. Kommandoen lager et diagram hvor du kan se median, kvartiler og mulige "outliers" (observasjoner som adskiller seg sterkt fra resten av observasjonene — noe som kan skyldes feil). BOXPlot er en grei kommando å bruke *før* du går igang med mer "avanserte" analyser, særlig hvis den benyttes sammen med BY-kommandoen. Vi skal ta for oss noen eksempler:

Vi har en kolonne kalt "Vekt", som inneholder kroppsvekten til endel personer av forskjellig kjønn. Vi begynner med å lage et BOXPlot for denne kolonnen:

```
MTB > BOXPlot 'Vekt'
```

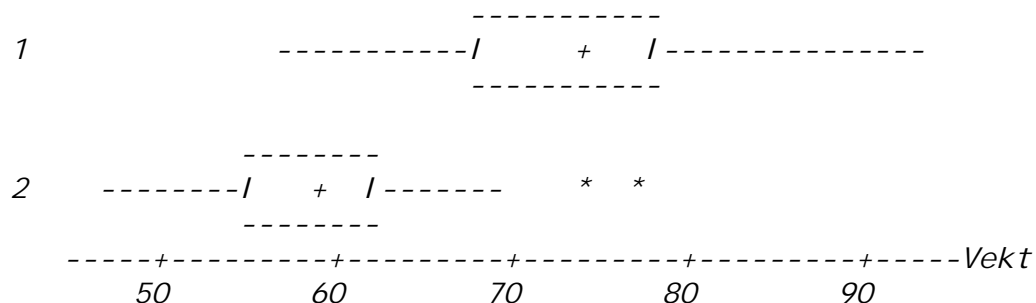


Dette ser jo i utgangspunktet ganske greit ut: Medianen ligger omkring 70 kilo, og observasjonene spenner fra 46 til 93 kilo. "Boksen" i diagrammet viser hvilke data som ligger innenfor øvre og nedre kvartil.

Vi har imidlertid en mistanke om at dette bildet er en smule upresist — at det kan være store forskjeller på kjønnene. Vi lager derfor et BOXPlot hvor vi ser på kjønnene hver for seg ved hjelp av subkommandoen BY (jfr. HISTogram-kommandoen):

```
MTB > BOXPlot 'Vekt';
SUBC> BY 'Kjonn'.
```

Kjonn



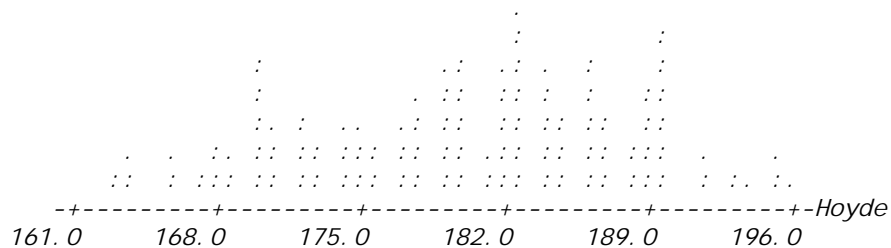
Vi ser her at det klart er forskjell mellom kjønnene. Medianen for menn ligger på ca. 74 kg, for

kvinner ca. 59. Vi ser også at det blant kvinnene er observert 2 "outliers" (ekstremverdier); disse er markert med med stjerner (*). Dette vil si at disse verdiene bør du kanskje se nærmere på, fordi de avviker sterkt fra de andre observasjonene. Dette trenger ikke nødvendigvis være feil (det kan jo godt tenkes at det er noen "tykksaker" blant de observerte kvinnene), men det er i alle fall godt håndverk å sjekke at det ikke er skrive- eller punchefeil.

5.14 DOTPLOT (punkthistogrammer)

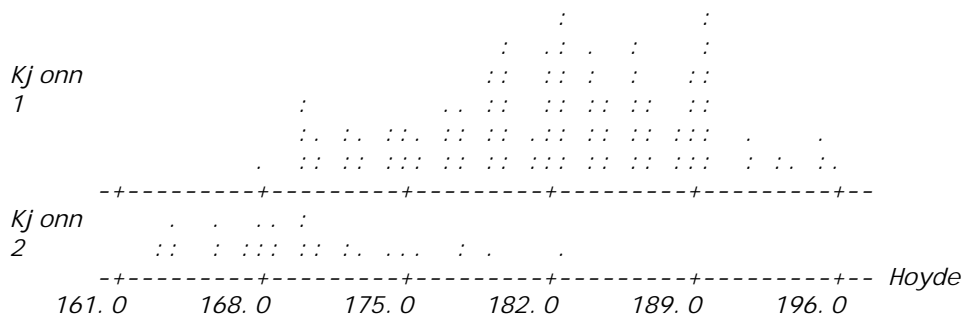
DOTPlot er en form for histogram — med vertikale søyler og punktum'er i stedet for asterisk'er (*) som symbol. Et eksempel:

```
MTB > DOTPlot C2
```



Også DOTPlot kan ta subkommandoen BY (se HELP DOTPlot for flere subkommandoer):

```
MTB > DOTPlot c2;
SUBC> BY c1.
```



Hvis du plottes to eller flere kolonner mot hverandre med DOTPlot eller BOXPlot-kommandoen, bør du passe på å benytte subkommandoen SAME. Den sørger for at tallinjen blir lik for alle kolonnene — ellers kan det bli vanskelig å sammenligne kolonnene rent visuelt. Et eksempel på bruk av SAME er vist under MANN-Whitney-kommandoen (side 100).

5.15 STEM-and-leaf diagram

Et STEM-and-leaf diagram er et slags histogram hvor man i stedet for et symbol (som f.eks. "**") benytter verdiene (tallene). Dette gjør det mulig å lese av de enkelte verdier, i tillegg til at man får en brukbar oversikt over fordelingen av et datasett.

Et eksempel:

MTB > STEM-and-Leaf ' Skonr. '

Stem-and-Leaf of Skonr. N = 177
Leaf Unit = 0.10

```

  1  36 5
 10  37 000000055
 25  38 000000000000055
 33  39 00000000
 43  40 0000000000
 65  41 0000000000000000055555
(31) 42 00000000000000000000000555555
 81  43 000000000000000000000000000000000555555555
 36  44 000000000000000005
 17  45 00000000000000
  4  46 0000
```

Dette første kolonnen i dette diagrammet viser det kumulative antallet observasjoner mot medianen (dvs. summen av antallet "mindre enn medianen" og "større enn medianen"). Den neste kolonnen viser hver "stamme" (her: de hele skonumrene). Deretter kommer "bladene" (her: tideler av skonumre).

Den "stammen" som inneholder medianen til dataene er satt innenfor parenteser. Av diagrammet ser vi f.eks. at det var 10 som hadde skonummer mindre enn 38; av disse hadde 2 skonummer 37.5, 7 hadde 37 og 1 hadde 36.5.

Lineær regresjonsanalyse

Dette kapittelet og de to følgende handler om regresjonsanalyse og beslektede statistiske metoder. Formatet er imidlertid anderledes enn resten av manualen: i stedet for kun å beskrive kommandoene og vise eksempler, forsøker jeg å gå mer i dybden. Grunnen til dette er at regresjonsanalyse er en av de viktigste metodene vi har i statistikk, samtidig som den er en av de "farligste"; jeg vil påstå (uten noe som helst statistisk signifikant grunnlag) at gal bruk av regresjonsanalyse er ansvarlig for flere feil prediksjoner enn alle andre statistiske teknikker tilsammen (muligens med konkurranse av faktoranalyse).

Formatet som følger: i dette kapittelet gjennomgår jeg regresjonsanalyse ved hjelp av Minitab og noen enkle eksempler. I det neste kapittelet, "Kontroll av regresjonsanalyse", gjennomgås de mest vanlige feilene spirende statistikere gjerne begår (for all del, de erfarne også), samt hva man kan gjøre for å unngå dem. Det siste kapittelet, "Avansert regresjonsanalyse" handler om endel triks man kan benytte seg av for å få frem mer kompliserte regresjoner.

Lineær regresjonsanalyse vil si at man søker å avdekke om det er noen sammenheng mellom to

$$Y = \beta_0 + \sum_{t=1}^n \beta_t x_t + \varepsilon$$

variable: den *avhengige* variabelen, som ofte kalles Y, og en (enkel regresjonsanalyse) eller flere (multippel regresjonsanalyse) uavhengige variable eller *prediktorer*, som benevnes X (hvis det er fler enn én, X_1, X_2 osv.). Uttrykt i en formel, er forholdet mellom Y og prediktorene

hvor

Y	er den avhengige variable
β_0	er et konstantledd (dvs. den verdien Y vil ha hvis alle X er 0)
β_t	er stigningstallet for variabelen x_t , dvs. hvor mye Y endrer seg hvis x_t endres.
ε	residualen, dvs. et uttrykk for tilfeldig variasjon i tillegg til "effekten" av prediktorene

I lineær regresjonsanalyse forutsetter man at sammenhengen mellom variablene er lineær, og en

$$\hat{Y} = b_0 + \sum_{t=1}^n b_t x_t$$

fremgangsmåte kalt *minste kvadraters metode* benyttes for å finne de ulike β -verdiene. Resultatet av en regresjonsanalyse er en formel for en *regresjonslinje*, hvor den avhengige variabelen "Y hatt" er en funksjon av regressorene

hvor $b_{0,t}$ er *regresjonskoeffisienter*, dvs. estimater for β -verdier. Andre resultater av regresjonsanalyse er en T-verdi for hver regresjonskoeffisient, som sier noe om hvor godt estimatet av regresjonskoeffisienten er i forhold til den ekte β -verdien, og r^2 , (også kalt determinasjonskoeffisienten, som sier noe om hvor mye av variasjonen i Y som kan forklares som

effekter av variasjoner i prediktorene.

Forvirret? Det hjelper å se på noen eksempler

6.1 Enkel lineær regresjon

Regresjonsanalyse utføres i Minitab med kommandoen REGRESS. Du kan utføre lineær regresjon med en eller flere prediktorer. Den generelle formen for REGRESS-kommandoen er

```
REGRESS Y-kolonne n x-kolonne(r) [st.res.kolonne [Y-hatt-kol.]]
```

hvor n er antall prediktorer (dvs. antall x-kolonner). Du *må* skrive n, selv om den er 1. Hvis du føyer til en ekstra kolonne i tillegg til de oppgitte x-kolonnen, blir de standardiserte residualene lagret i den. Hvis du føyer til enda en, lagres de predikerte Y-verdiene (̂) i den.

La oss se på et banalt eksempel:

Sherlock Holmes benyttet en persons skrittlengde for å avgjøre hvor høy han eller hun var. Vi har ikke tall for studenters skrittlengde — men vi har skonummerne. Er det mulig å predikere, ut fra oppgitt skonummer, hvor høy en student er?

Vi gir denne kommandoen i Minitab (subkommandoen RESIDUALS får vi bruk for ved senere residualanalyse, C28 og C29 vil inneholde henholdsvis standardavvikene til hver predikerte verdi og den predikerte verdien):

```
MTB > REGRESS 'Hoyde' 1 'Skonr.' c28 c29;  
SUBC> RESIDUALS c30.
```

```
The regression equation is  
Hoyde = 63.2 + 2.80 Skonr.
```

Predictor	Coef	Stdev	t-ratio	p
Constant	63.155	6.214	10.16	0.000
Skonr.	2.7956	0.1482	18.86	0.000

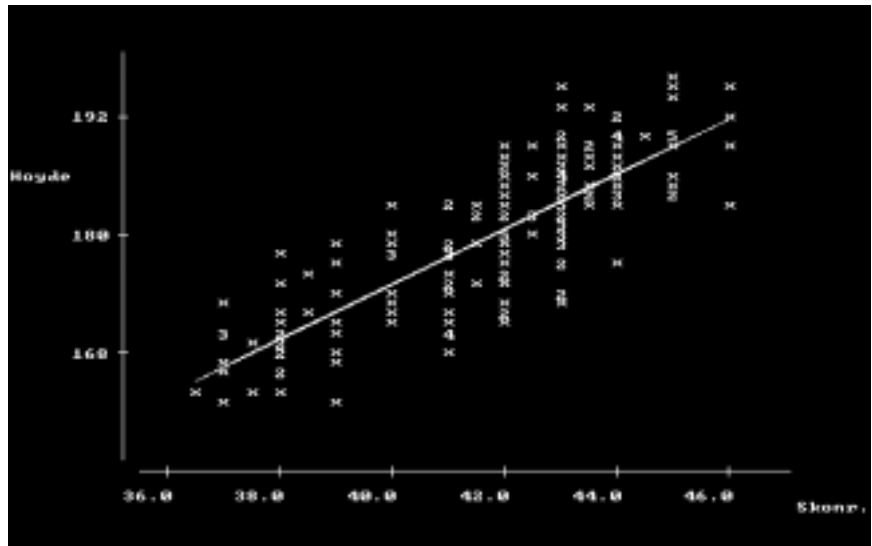
```
s = 4.500      R-sq = 67.0%      R-sq(adj) = 66.8%
```

Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	1	7203.1	7203.1	355.63	0.000
Error	175	3544.5	20.3		
Total	176	10747.6			

```
Unusual Observations  
Obs. Skonr. Hoyde      Fit Stdev. Fit Residual St. Resid  
36  43.0 174.000 183.364  0.379  -9.364  -2.09R  
.  
.  
.  
.  
.  
.  
.  
.  
.  
.  
169 36.5 164.000 165.193  0.863  -1.193  -0.27X
```

R denotes an obs. with a large st. resid.
X denotes an obs. whose X value gives it large influence.

Som vi ser, skrives først regresjonsligningen ut (The regression equation is...). Hvis det hadde vært noen manglende observasjoner i en av kolonnene, ville vi fått beskjed om det her — Minitab kutter ut slike observasjoner fra regresjonsanalysen. Så kommer en tabell med standardavvik og t-verdier for konstantleddet og stigningstallet til regresjonslinjen.



Deretter følger s (estimert standardavvik til linjen), R^2 ("forklart" variasjon/total variasjon, også kalt determinasjonskoeffisienten). R^2 -adjusted er determinasjonskoeffisienten justert for antall frihetsgrader. Denne vil bli mindre i forhold til R^2 hvis vi øker antallet prediktorer.

Så følger en tabell som viser detaljer om variansen. SS er kvadratsummen av variasjonene rundt linjen. $SS(\text{Regression})/SS(\text{Total})$ gir R^2 .

Til slutt kommer en tabell over "uvanlige" observasjoner — dvs. observasjoner som adskiller seg fra de "vanlige" på en av to måter: enten ved at residualen er stor (stor forskjell mellom predikert verdi (\hat{Y}) og observert verdi (Y)), eller at X -verdien enten er stor eller liten, og derfor bør sjekkes for å se om det er mulige "outliers" (ekstremverdier).

Det alt dette betyr, er at man kan estimere en persons høyde til å være 63.2 cm + 2.80 ganger skonummeret¹³. Imidlertid kan man ikke regne med at mer enn ca. to tredjedeler av variasjonen i høyde kan forklares ut fra skonummeret — hvilket kan tyde på at det er begrenset sammenheng mellom høyde og skonummer, og at Sherlock Holmes kanskje skulle forsøke å finne flere holdepunkter.

6.2 Multippel regresjonsanalyse

zzzz skikkelig eksempel (NSF)

Multippel lineær regresjonsanalyse vil si regresjonsanalyse med flere enn én prediktor. Kommandoen for **multippel lineær regresjon** er

```
REGRESS Y-kolonne n x-kolonne(r) [st.res.kolonne [Y-hatt-kolonne]]
```

hvor n er antall prediktorer (dvs. antall x -kolonner). Dette tilsvarer kommandoen for enkel lineær regresjon.

¹³ Dette er selvfølgelig et punktestimert — for å beregne et konfidensintervall rundt dette estimatet kan du benytte REGRESS-kommandoens subkommando PREDICT (s.d.).

6.2.1 Mer om REGRes-kommandoen

REGRes-kommandoen har en rekke subkommandoer, noen av dem er:

- RESiduals, som lar deg spesifisere en kolonne for lagring av residualene (dette er greit ved senere "sjekk" av residualene, se eget punkt om residualanalyse).
- NOCOntant, som lar deg foreta regresjonsanalyse uten bruk av konstantledd (dvs. uten skjæringspunkt med Y-aksen). Dette brukes i tilfeller hvor du på forhånd kan si at det ikke skal være noe konstantledd — f.eks. når du ser på utviklingen av et marked over tid, hvor du vet at verdien ved markedets begynnelse var 0.
- DW, som beregner og skriver ut Durbin-Watson tallet for autokorrelasjon.
- PREDict, som lar deg oppgi X-verdier (enten som konstanter eller som kolonner) som det skal prediseres Y-verdier for. Fordellen med ikke å gjøre dette manuelt er (bortsett fra at man slipper å regne ut resultatet) er at man får et standardavvik til estimatet oppgitt, og dermed kan kjøpt og greit regne ut et konfidensintervall.

Hvor mye opplysninger som skrives ut fra hver regresjon kan reguleres ved hjelp av BRIEF-kommandoen (s.d.)

6.3 Korrelasjon

Korrelasjonskoeffisienten mellom to datasett er et mål på i hvor stor grad de samvarierer — uten at man gjør noe forsøk på å finne ut om i hvilken grad man kan predikere en variabel ved hjelp av en annen. I Minitab kan du beregne korrelasjonskoeffisienten mellom kolonner ved kommandoen

```
CORRelate kolonne1 kolonne2 . . . kolonneN [lagres i matrise M]
```

La oss ta et eksempel; vi ønsker å beregne korrelasjonen mellom kjønn, høyde, vekt og skonummer, som ligger i kolonnene C1-C4. Vi gir denne kommandoen:

```
MTB > CORRelate c1-c4
```

	<i>Kjonn</i>	<i>Hoyde</i>	<i>Vekt</i>
<i>Hoyde</i>	-0.735		
<i>Vekt</i>	-0.687	0.827	
<i>Skonr.</i>	-0.838	0.856	0.800

og får en korrelasjonsmatrise for de aktuelle variablene. Som vi ser er det høy grad av samvariasjon mellom dem (det burde jo ikke forundre noen).

Kontroll av regresjonsanalyse

There are three kinds of lies:

*lies,
damned lies,
and statistics.*

Benjamin Disraeli

Dette kapitlet er uten tvil det viktigste i hele denne manualen, fordi det handler om hvordan man ikke skal benytte statistikk. Statistikk er skummelt fordi det er så overbevisende: man glemmer fort hvor mye antagelser som ligger bak de fleste analyser. For eksempel er det svært vanlig å benytte 5% signifikansnivå — dette betyr i praksis at man garantert tar feil 1 gang av 20. Det er imidlertid ikke bare statistikken som kan føre en på ville veier — det finnes en mengde andre måter å trekke gale konklusjoner eller lage rare teorier på. De amerikanske statistikerne Thomas D. Cook og Donald T. Campbell skrev i 1977 en bok¹⁴ som har blitt en slags bibel om hvordan man skal unngå å lage gale teorier på grunnlag av statistikk. De benytter *validitet* som kriterium for hvor god en konklusjon er, og deler inn validitet i fire typer, med et spørsmål å stille til hver:

Type validitet	Spørsmål	Ting å passe på
<i>Statistisk konklusjonsvaliditet</i> ("statistical conclusion validity")	Er det en samvariasjon mellom de to (eller flere) variablene?	- heteroscedastisitet - multikolaritet - normalitet - autokorrelasjon
<i>Intern validitet</i> ("internal validity")	Gitt at det er samvariasjon, er det en plausibel sammenheng?	- kausalitet - fisketurer
<i>Fortolkningsvaliditet</i> ("construct validity")	Gitt at sammenhengen er plausibel, hva er de presise årsaksforhold (direkte påvirkning eller mellomliggende variable)?	- mellomliggende variable - forskjellen mellom nødvendighet og tilstrekkelighet
<i>Ekstern validitet</i> ("external validity")	Gitt de tre ovenfor, i hvilken grad kan man generalisere konklusjonene fra det foreliggende datamaterialet?	- er dette et spesialtilfelle?

I de neste punktene vil jeg ta for meg hver enkelt av disse typene validitet, og hvordan man kan sørge for å være gardert mot å gå i vannet for hver enkelt av dem.

¹⁴*Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago: Rand-McNally. Oppdelingen i fire typer validitet er en videreutvikling av Campbell & Stanley's (1963) inndeling av validitet i intern (omfatter de to første typene ovenfor) og ekstern (de to siste) validitet. Intern validitet omhandler en teori eller konklusjon som sådan, ekstern validitet graden av hvorvidt konklusjonen kan ha konsekvenser utenom det område undersøkelsen er foretatt innenfor.

7.1 Statistisk konklusjonsvaliditet

Statistisk konklusjonsvaliditet er stort sett et spørsmål om å foreta endel tester for å se hvor bra en regresjonsanalyse er: dette vil si at man ikke trenger å vurdere hva som ligger bak tallene i denne omgang. I hovedsak må man sjekke hvorvidt alle forutsetninger for bruk av regresjonsanalyse er oppfylt. Regresjonsanalyse forutsetter

Ingen *multikolaritet*: at det ikke er for høy grad av samvariasjon mellom prediktorene.

Normalitet: dvs at residualene (forskjellene mellom punkter på regresjonslinjen og faktiske observasjoner) er kommet fra en populasjon som er normalfordelt med 0 som gjennomsnitt (dvs. at de er tilfeldige)

Homoscedastisitet: at det ikke er forskjell i spredningen av residualene langs x-aksen (heteroscedastisitet)

Ingen *autokorrelasjon*: hver verdi av X skal være uavhengig av den verdien som kom før den¹⁵.

7.1.1 Multikolaritet

Multikolaritet betyr at prediktorene er høyt korrelert med hverandre, og er altså bare et problem i multippel regresjonsanalyse. Hvis man har multikolaritet, vil regresjonsanalyse gi en meget ustabil modell, noe man ofte ser ved at regresjonskoeffisientene varierer mye hvis man legger til nye prediktorer (Lapin, 1987). Man sjekker for multikolaritet ved å kjøre en korrelasjonsanalyse (CORRelation-kommandoen) for alle prediktorene. Som en tommelfingerregel kan man si at for de fleste anvendelser av regresjonsanalyse kan man se bort fra multikolaritet hvis ingen av korrelasjonskoeffisientene mellom hvert prediktor-par (dvs. ingen av cellene i korrelasjonsmatrisen) er større enn ± 0.8 ($\pm 80\%$) (Berk, 1983, side 535).

Minitab har en "automatisk" kontrollfunksjon for multikolaritet: Hvis korrelasjonen mellom en prediktor og en annen er høy (innbyrdes R^2 er høyere enn 99%) skrives denne feilmeldingen ut:

NOTE xx is highly correlated with other predictor variables

Hvis korrelasjonen er meget høy ($R^2 > 99.99\%$) skrives denne feilmeldingen ut:

xx is highly correlated with other predictor variables
xx has been omitted from the equation

Prediktoren blir dermed ikke tatt med (et eksempel på dette er vist under punktet om regresjon med dummy-variable).

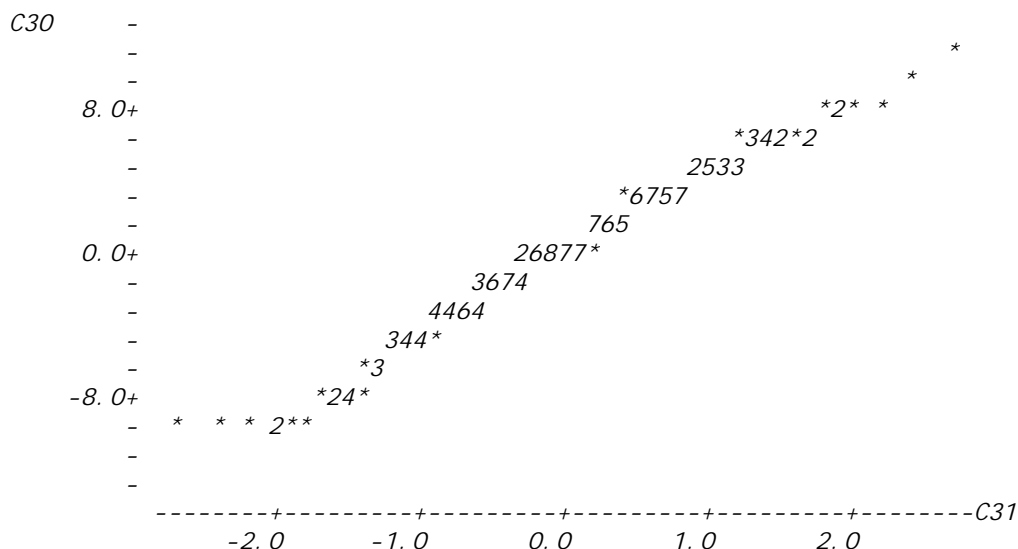
7.1.2 Normalitet

Normalitet vil si at residualene er normalfordelt for alle verdier av prediktoren(e). Regresjonsanalyse er ganske robust overfor brudd på denne forutsetningen (Berenson & Levine, 1979 s. 493), og som regel kan du se bort fra dette når antallet observasjoner er over 100 (Berk, 1983, s.510). Det finnes en

¹⁵ Dette er mest aktuelt ved tidsserieanalyser, og vil ikke bli behandlet her.

enkel metode for å sjekke normalitet, kalt residualanalyse. Ved å bruke subkommandoen RESiduals kan du lagre residualene inn i en egen kolonne (vist under eksemplet om enkel regresjon). Deretter kan du analysere dem ved hjelp av denne teknikken:

```
MTB > NSCOres c30 c31
MTB > PLOT c30 c31
```



Hvis residualene kommer fra en normalfordelt populasjon med populasjonsgjennomsnitt (μ) = 0, vil denne teknikken gi en (tilnærmet) rett linje. Her ser vi at det muligens kan være grunn til å være forsiktig for de lavere skonomre.

NSCOres-kommandoen beregner for hver verdi i kildekolonnen den tilsvarende sannsynlige verdi hvis det hadde vært trukket fra en normalfordelt populasjon med gjennomsnitt = 0 og standardavvik = 1. Antall trekninger tilsvarer antall observasjoner i kildekolonnen.

Korrelasjonen mellom residualene og deres "normal scores" skal være høy (dvs. nærme seg 1):

```
MTB > CORRelati on c30 c31
```

```
Correlation of C30 and C31 = 0.996
```

Grenseverdier, basert på signifikansnivåer på hhv. 10, 5 og 1% er vist i tabellen¹⁶ på side 86. Hvis korrelasjonskoeffisienten er lavere enn disse verdiene for det aktuelle antall (N) og signifikansnivå (α), er

N	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.01$
4	0.8951	0.8734	0.8318
5	0.9033	0.8804	0.8320
10	0.9347	0.9180	0.8804
15	0.9506	0.9383	0.9110
20	0.9600	0.9503	0.9290
25	0.9662	0.9582	0.9408
30	0.9707	0.9639	0.9490
40	0.9767	0.9715	0.9597
50	0.9807	0.9764	0.9664
60	0.9835	0.9799	0.9710
75	0.9865	0.9835	0.9757
80	0.	0.	0.
100	0.9898	0.9874	0.9804
150	0.9935	0.9918	0.9853
200	0.9956	0.9942	0.9878
300	0.9980	0.9968	0.9904
400	0.9993	0.9983	0.9917
600	0.	0.	0.
1000	0.	0.	0.

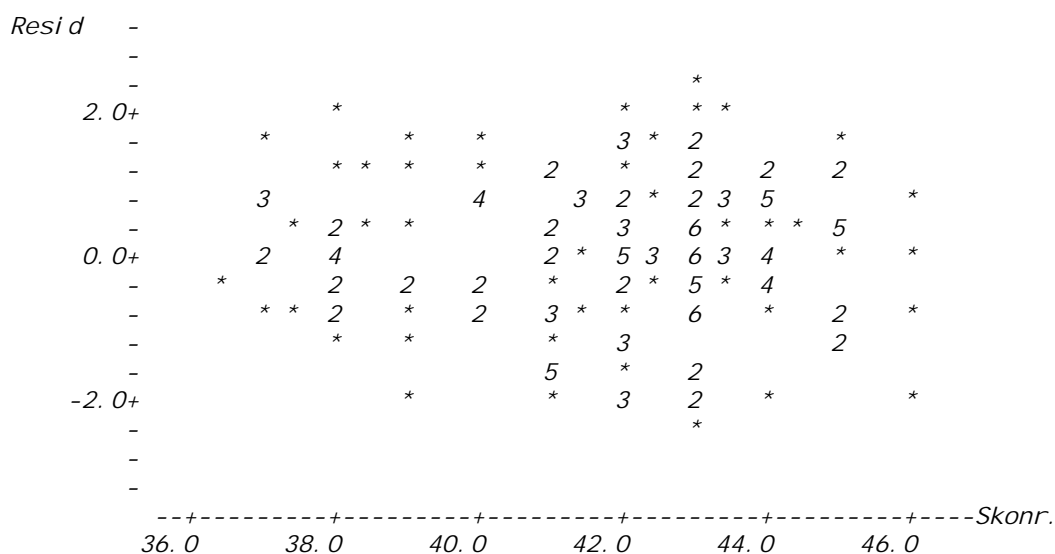
¹⁶ Disse verdiene er hentet fra Minitab's Release 8 Manual. Noen verdier er beregnet etter tilnæringsmåte beskrevet i Ryan & Joiner (1976).

det grunn til å sette spørsmålstegn ved om det kan være en lineær sammenheng mellom variablene.

7.1.3 Heteroscedastisitet

For å kontrollere for heteroscedastisitet (dvs. at spredningen varierer langs regresjonslinjen) kan du PLOTte residualene mot hver av prediktorene. Vi tar utgangspunkt i eksemplet med forholdet mellom høyde og skonommer, med residualene lagt inn i kolonne 29, og de prediserte verdiene (Y-hatt) i kolonne 30.

```
MTB > NAME C29 'Resid' C30 'Y-hatt'
MTB > PLOT 'Resid' 'Skonr.'
```



Av dette plottet ser vi at det kan være tegn til at residualene er mer spredd rundt de høyere skonommerne — men det er ikke snakk om store forskjeller. Allikevel kan det være grunn til å se om det ikke finnes andre variable enn skonommer for å finne en person's høyde, noe vi kommer til å gjøre under punktet om regresjon med dummy-variable (pkt. ?, side 90).

Hvis man har heteroscedastisitet, vil gjerne regresjonskoeffisientene være riktig estimert, men med for store standardavvik (dvs. for dårlige T-verdier), hvilket kan medvirke til feil av type II (at man forkaster en god hypotese) (Berk, 1983, s. 516-517). Med andre ord, om man har svært gode T-verdier, trenger man ikke bekymre seg for mye om heteroscedastisitet (som i eksemplet ovenfor).

7.2 Intern validitet

Regresjonsanalyse forsøker å påvise samvariasjon mellom variable — og hvor sterk samvariasjonen er. Den har to "feller" som mange håpefulle studenter (og etterhvert endel forskere) går i: at man tester data for å finne teori, og at man glemmer å sjekke forutsetninger.

La oss se på dette med teori først: Regresjonsanalyse påviser *samvariasjon* - ikke nødvendigvis sammenheng (årsaksforhold). Følgende historie skulle gjøre det hele så meget klarere:

I et middagsselskap ble det diskutert hvor usunt det egentlig var å gå med sko. En av festdeltagerne (som hadde forsynt seg godt av drikkevarene) bidro med følgende observasjon: "Det där e' aldeles riktig; varje gång jag vaknar med skor på mig har jag huvudvärk!

(Albert Engström)

God statistisk metode (generell vitenskaplig metode for den saks skyld) går ut på at man setter seg ned og tenker over et problem, finner frem til en mulig løsning (teori), finner ut hvilke data man trenger for å teste teorien, tester dataene, og til slutt rapporterer hvordan dataene stemte med teorien. Når man måtte bruke papir og blyant eller regnestav var det ikke noe problem med denne rekkefølgen: hvis man tenkte hardt og lenge på teorien og hvilke data man trengte, ble mengden beregninger som måtte foretas kraftig redusert. Etter at man fikk datamaskiner og statistikkprogrammer, begynte en annen metode å dukke opp: Man samlet sammen en masse data, og foretok en masse beregninger nærmest tilfeldig. I noen tilfelle kan dette resultere i at man finner noen nye teorier som man ikke hadde tenkt på før. Mer vanlig er det at man kommer frem til forklaringer som nok ser bra ut på papiret (dvs. at statistiske parametre som T-verdier og forklart varians er bra) men som ikke holder vann hvis man forsøker å bruke dem til noe praktisk zzzz. (ta med poeng om at dette er OK ved explorativ forskning, men ikke innenfor et kurs i elementær statistikk).

zzzz utbygges Imidlertid, en advarsel er på sin plass: Det er fristende å benytte Minitab's (og for den saks skyld andre statistikkprogrammer) til å "fiske" etter samvariasjon: man kjører en masse regresjoner, finner en modell med høy forklart varians (R^2), og kaster ut eller inkluderer prediktorer (uavhengige variable) til man sitter igjen med noe som ser ut som om det forklarer det man er ute etter. Dette er en *meget* uheldig måte å benytte statistikk på — man kan kanskje oppnå en elegant modell, men hvis man forsøker å benytte den til prediksjon (dvs. til noe praktisk) vil man som regel havne i vanskeligheter. Den rette fremgangsmåten (som i all vitenskaplig metode) er å sette seg ned og forsøke å gjøre seg opp en formening om hvordan ting bør henge sammen — og så teste disse hypotesene. Spørsmålet om man skal inkludere enkelte variabler i analysen eller ikke bør avgjøres ut fra en oppfatning om variablene har relevans eller ikke — ikke ut fra tilfeldig samvariasjon.

- zzzz idé: splitt "gutt" og "jente" i to regresjoner, hvis hvordan det tar seg ut med en liten graf (med to linjer)

7.3 Fortolkningsvaliditet

zzzz

7.4 Ekstern validitet

- zzzz del opp dataene i to deler (med RANDom-kommandoen) og sjekk om du kan predisere den ene halvparten basert på den andre.

7.5 Nyttige triks

Ikke-lineær regresjonsanalyse

De to foregående kapitlene har beskrevet *lineær* regresjon: forholdet mellom den avhengige variabelen og prediktorene antas å følge en rett linje. zzzz

- også problem hvis dataene er diskrete (ikke kontinuerlige)
- zzzz gamma funksjon, muligens simulering av LOGIT/PROBIT (se KeepingTAB)

7.5.1 Regresjon med indikatorer (dummy-variable)

Bruken av indikatorer (også kalt dummy-variable) er en måte å foreta regresjonanalyse hvor prediktorene er klassifikasjoner i stedet for kontinuerlige variable. Indikatorene er enten 0 eller 1. Du kan bruke Minitabs INDicator-kommando for å lage indikatorkolonner fra en kolonne med diskrete data.

Først den generelle formen på INDicator-kommandoen:

INDicator kildekolonne målkolonner

Her må det være like mange målkolonner som det er forskjellige typer observasjoner i kildekolonnen.

Forvirret? La oss ta et eksempel:

Hvis vi har en kolonne 'Kjonn' (1=Mann, 2=Kvinne), kan vi gi kommandoen

```
INDicator 'Kjonn' C20 C21
```

Deretter navngir vi C20 og C21 (som nå inneholder dummy-variable) slik:

```
NAME C20 'Mann' C21 'Kvinne'
```

Resultatet:

Kjonn	Mann	Kvinne
2	0	1
1	1	0
1	1	0
2	0	1
1	1	0
2	0	1
1	1	0

1 1 0

Deretter kan du kjøre enkel eller multipel regresjonsanalyse mot indikatorkolonnene, f.eks. for å se om vi kan si noe om en persons skomnummer på grunnlag av om hvilket kjønn personen er av. Vi tar med kolonnen 'Royking' også — den inneholder også en indikatorvariabel.:

```
MTB > REGress 'Skonr.' 3 'Mann' 'Kvinne' 'Royking'
```

```
* Kvinne is highly correlated with other X variables  
* Kvinne has been removed from the equation
```

```
The regression equation is  
Skonr. = 37.9 + 4.66 Mann + 0.356 Royking
```

```
173 cases used 4 cases contain missing values
```

Predictor	Coef	Stdev	t-ratio	p
Constant	37.9335	0.2739	138.48	0.000
Mann	4.6570	0.2850	16.34	0.000
Royking	0.3561	0.2505	1.42	0.157

```
s = 1.439 R-sq = 61.2% R-sq(adj) = 60.7%
```

```
Analysis of Variance
```

SOURCE	DF	SS	MS	F	p
Regression	2	555.01	277.50	134.07	0.000
Error	170	351.86	2.07		
Total	172	906.87			

SOURCE	DF	SEQ SS
Mann	1	550.83
Royking	1	4.18

```
Unusual Observations
```

Obs.	Mann	Skonr.	Fit	Stdev. Fit	Residual	St. Resid
26	1.00	46.000	42.590	0.135	3.410	2.38R

```
. . . . (utskriften forkortet)
```

153	1.00	46.000	42.590	0.135	3.410	2.38R
-----	------	--------	--------	-------	-------	-------

```
R denotes an obs. with a large st. resid.
```

Som vi ser, kutter Minitab ut 'Kvinne'-kolonnen fordi den er høyt korrelert (korrelasjonsfaktor = -1) med kolonnen for 'Mann' (med andre ord — et ekstremt eksempel på *multikolinearitet* (s.d.). Dette betyr i og for seg ingenting — siden 0 i "mannskolonnen" betyr at personen er kvinne. Regresjonsligningen blir således et mål for gjennomsnittet av kvinnenenes skomnummer — med det tillegg som må gjøres hvis det er en mann — og hvis personen røyker. Som vi ser, gir kolonnen om røyking liten tilleggsinformasjon.

7.5.2 Transformasjon av variable

zzzz et annet moment (som også gjør seg gjeldende ved enkel regresjon) er at for å få et operasjonelt ("brukbart i praksis") tall for sammenhengen mellom en variable og en prediktor må ikke den tallmessige variasjonen for prediktoren være for liten.

Et eksempel (tatt fra Minitab Reference Manual) kan være at man bruker årstall som prediktor. Man bør da ikke bruke f.eks. årstall som f.eks. 1970-1989, men heller 0 - 19 (dvs. trekke fra en konstant, i dette tilfelle 1900, fra prediktorkolonnen). Dette har ikke noe å si for kvaliteten av regresjonsanalysen — men man får mer "lesbare" tall. Du kan se at dette er tilfelle hvis tallet foran prediktoren er svært nær 0, samtidig som R^2 er relativt høy.

Minitabs regresjonskommando bruker "minste kvadraters metode" for å beregne lineær regresjon. Hvis du har en mistanke om at sammenhengen er ikke-lineær, kan du prøve seg med en 3. grads ligning etter denne teknikken:

```
MTB > LET C10 = 'Vekt' ** 2
MTB > LET C11 = 'Vekt' ** 3
MTB > NAME C10 'Vekt**2' C11 'Vekt**3'
```

Først opphøyer du dataene for den uavhengige variabelen i annen og tredje potens. Deretter kan du kjøre Minitabs REGRes-kommando med 3 prediktorer:

```
MTB > REGR 'Hoyde' 3 'Vekt' 'Vekt**2' 'Vekt**3'
```

(utskriften fra denne kommandoen er ikke tatt med her — den gir ingen mening for disse dataene).

7.5.3 Gamma-funksjonen

Gamma-funksjonen er en funksjon som uttrykker et forhold mellom to variable som følger en linje slik du ser den i figur zzzz. Y øker med stigende X, først langsomt, så fortere, for så å flate ut. Denne funksjonen ser man ofte i praksis: f.eks. beskriver den mange produkters livssyklus, dvs. salgstall over tid.

Ligningen for denne funksjonen er

$$\log_e(\hat{Y}) = b_0 + b_{\text{yearexp}} \text{yearexp} + b_{l(\text{year})} l(\text{year}) + b_{l(\text{realh})} l(\text{realh}) + \sum_{\text{dummy } t=1}^n b_{\text{dummy } t} X_{\text{dummy } t} + \varepsilon$$

zzzz utled funksjonen skikkelig.

7.5.4 Log-log regresjon

zzzz To build models of relative rather than nominal effects, one may use OLS with the natural logarithm of Y as the dependent variable and the natural logarithms of $X_{1..n}$ as independent variables (Hanushek and Jackson, 1977, p.98). This has the added convenience of making the coefficients produced the elasticities of the relationship between the variables, giving the researcher the effect of a percentage shift in any X on the dependent variable (given that it is a small change, and *ceteris paribus*, i.e., all the other predictors are held constant). The results of the log-log approach are more

generalizable in modeling economical relationships, since outside effects as inflation can be ignored when using the model to make inferences over more than one time period. The nominal model has the virtue of being more easily communicated to lay readers; both the results and the equations are more intuitive.

7.6 To skumle kommandoer

De to kommandoene som gjennomgås her, er ansvarlig for en rekke idiotiske statistiske konklusjoner — fordi de oppmuntrer til "fisking" etter data i stedet for skikkelig hypotesetesting. Allikevel vil jeg ta dem med, av den grunn at det er bedre at de er i hvert fall nogenlunde forklart her, enn at folk skal bruke dem kritikkløst etter å ha funnet dem i Minitab-manualen eller hjelpetekstene.

7.6.1 Å finne de beste prediktorene

BREGress-kommandoen (BREGress står for *best subset regression*) er en kommando som hurtig går igjennom alle mulig kombinasjoner av prediktorer mot en avhengig variabel og forteller deg hvem av dem som egner seg best til videre analyse.

Den generelle formen for kommandoen er

BREGress *Y-kolonne x-kolonner*

La oss ta et eksempel:

Vi ønsker å finne ut hvilken eller hvilke variabler som kan gi størst forklaring på en persons høyde. Vi gir kommandoen

MTB > **BREGress c2 c1 c3-c5 c6**

Best Subsets Regression of Hoyde

173 cases used 4 cases contain missing values.

<i>Vars</i>	<i>R-sq</i>	<i>Adj. R-sq</i>	<i>C-p</i>	<i>s</i>	<i>Step AIC</i>
1	67.5	67.4	38.7	4.4489	X
1	66.2	66.0	47.4	4.5418	X

2	73.2	72.9	4.6	4.0558	X X
2	67.7	67.4	39.4	4.4488	X X
3	73.5	73.1	4.5	4.0425	X X X
3	73.5	73.0	4.5	4.0434	X X X
4	73.8	73.1	4.9	4.0365	X X X X
4	73.7	73.1	5.3	4.0401	X X X X
5	73.9	73.1	6.0	4.0371	X X X X X

Som vi ser, går BREG gjennom alle variablene, og skriver ut de to beste kombinasjonene for hvert subsett; dvs. de to beste prediktorene hvis du kun skal ha med en prediktor (i dette tilfelle Skonummer og Vekt, de to beste kombinasjonene av to prediktorer (i dette tilfelle Skonummer/Vekt og Skonummer/Sigarett) osv. Det kan kanskje virke litt underlig at kombinasjonen Skonummer/Sigarett skulle komme ut som nr. 2, i og med at det tidligere er vist at det for disse dataenes vedkommende ikke er sammenheng mellom kroppshøyde og antall sigaretter. Hvis du sjekker justert R^2 , vil du imidlertid se at den er lik for Skonummer og kombinasjonen Skonummer/Sigarett — variabelen Sigarett gir altså ingen forbedret prediksjon, og skal derfor ikke tas med.

Du kan regulere antallet utskrevne kombinasjoner med subkommandoen BEST (dvs. at hvis du ønsker de 3 beste kombinasjonene innen hvert subsett kan du gi subkommandoen BEST 3).

7.6.2 Trinnvis multippel regresjonsanalyse

STEPwise regresjon (som er et greit verktøy når du skal avgjøre hvilke prediktorer du skal bruke) gjøres i Minitab ved at du spesifiserer en avhengig variabel (dvs. en kolonne) og en rekke prediktorer. Minitab går deretter gjennom disse prediktorene og kommer opp med en liste over i hvilken grad disse prediktorene gir en bedre forklaring. Dette beregnes ved at Minitab går trinnvis frem: først beregnes R^2 for hver enkelt prediktor. Deretter velges den som gir størst R^2 ut. Så fortsetter prosedyren med de gjenværende prediktorene — basert på det **tillegg** til R^2 de vil gi etter at den første prediktoren er tatt med. Prosessen avbrytes når F-verdien kommer under 4.

Et raskt eksempel:

Vi sjekker 'Hoyde' mot de andre variablene (se tabell bakerst i manualen) for å se i hvilken rekkefølge vi bør foreta regresjonsanalyser med 'Høyde' som avhengig variabel.

```
MTB > STEPwise se C2 C1 C3-C6

STEPWISE REGRESSION OF Hoyde
ON 5 PREDICTORS, WITH N = 173
N(CASES WITH MISSING OBS.) = 4
N(ALL CASES) = 177

STEP          1          2
CONSTANT     63.52     89.52

Skonr.        2.79        1.60
T-RATIO      18.87        6.67

Vekt          0.334
```

<i>T-RATIO</i>		5.98
<i>S</i>	4.45	4.06
<i>R-SQ</i>	67.55	73.19

Som vi ser, er 'Skonr.' den prediktor som gir størst R^2 — deretter kommer 'Vekt', som står for den største økningen etter at 'Skonr.' har "virket". Prosessen stopper deretter, fordi F blir for liten (dvs. "neste F" vil bli under 4).

Det er mulig ved hjelp av subkommandoer å forandre metoden Minitab bruker for STEPwise regresjon. Du kan bl.a.

- benytte "backward elimination" STEPwise regresjon (starte med mengde prediktorer og ta bort en og en)(FENTER=0)
- benytte "forward selection" teknikk (som er en mer rigid utgave av standard-teknikken) (FREMmove=0)
- spesifisere grensen for at en prediktor skal fjernes/tas med (FENTER/FREMmove)
- spesifisere rekkefølge selv (FORCE)
- spesifisere om prediktorer skal tas med eller fjernes for hvert trinn (ENTER/REMOve)

Det henvises til Minitab Reference Manual for en nærmere beskrivelse av disse subkommandoene.

7.7 Diskriminantanalyse

Diskriminant-analyse benyttes hvis den avhengige variabel (Y) er av "kategoritypen" — mens prediktorene er kontinuerlige. Dette er ofte tilfelle innen f.eks. markedsanalyse, hvor man er interessert i å inndele befolkningsgrupper i markedssegmenter o.l. Diskriminantanalyse er ikke beskrevet i detalj her — kun vist ved et eksempel.

Den generelle formen for diskriminantanalyse i Minitab er

DISCriminant kategorikolonne prediktorkolonne(r)

Et eksempel:

Er det mulig å avgjøre en persons kjønn hvis du kun kjenner hans/hennes høyde og vekt?

Vi prøver dette:

```
MTB > DISCriminant c1 c2 c3
Linear Discriminant Analysis for Kjønn

Group      1      2
Count     145    32

Summary of Classification

Put into . . . . True Group. . . .
Group      1      2
1         131    2
```

2	14	30
Total N	145	32
N Correct	131	30
Proport.	0.903	0.937

N = 177 N Correct = 161 Prop. Correct = 0.910

Dette er den såkalte klassifikasjonsmatrisen, som angir hvor mange i hver gruppe Minitab klarte å klassifisere riktig (dvs. "bestemme kjønn på"). Deretter følger endel data om forskjellen mellom de to gruppene, samt til slutt.

Squared Distance Between Groups

	1	2
1	0.00000	5.08910
2	5.08910	0.00000

Linear Discriminant Function for Group

	1	2
Constant	-543.27	-504.66
Høyde	7.09	6.95
Vekt	-2.79	-2.99

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Sqrd Distnc	Probability
1 **	2	1	1	1.124	0.695
			2	2.775	0.305
2 **	1	2	1	2.9244	0.238
			2	0.5982	0.762
(. . . . utskriften forkortet)					
159 **	1	2	1	4.60003	0.095
			2	0.08344	0.905
176 **	1	2	1	4.9861	0.097
			2	0.5243	0.903

. . . . en liste over de observasjoner Minitab *ikke* klarte å klassifisere riktig — med klassifikasjonssannsynligheter for hver enkelt observasjon.

Som vi ser, klarte Minitab å klassifisere 161 av 177 personer riktig på grunnlag av deres høyde og vekt. Vær imidlertid oppmerksom på at det er høy grad av multikolinearitet mellom høyde og vekt.

Vi ser også at det var noe lettere å klassifisere jentene enn guttene (30 av 32 jenter riktig plassert, 131 av 143 gutter).

Ikke-parametriske tester

Det følgende er en kortfattet forklaring på hvordan du kan få Minitab til å foreta endel ikke-parametriske analyser av datasett. Ikke-parametriske tester forutsetter ikke at populasjonen er normalfordelt. Dette gjør at de er ofte brukt på små datasett.

7.8 Test av rekkefølge (RUNS)

RUNS-kommandoen gjennomfører en tosidig test av hvorvidt data i en kolonne ligger i tilfeldig rekkefølge. Den generelle formen på kommandoen er:

```
RUNS [x] kolonne(r)
```

hvor x er den verdien det testes "over og under"- dvs. grenseverdien for en "run". Hvis du oppgir noen x, bruker Minitab gjennomsnittet av kolonnen.

Et eksempel: Vi har en rekke med tall (0 og 1) som vi antar representerer myntkast (hvor 0 er "mynt" og 1 er "krone"). Rekken er lagret i kolonne C1 og ser slik ut:

```
0 1 1 1 0 1 0 0 0 1 1 0 1 0 0 0 1 1 1 1 0 0 1 1 1 0 0 1 0 0
```

Vi setter opp hypotesen H_0 : rekkefølgen er tilfeldig, og H_1 : den er ikke tilfeldig, og gir denne kommandoen i Minitab:

```
MTB > RUNS c1
```

```
C1
```

```
K = 0.5000
```

```
THE OBSERVED NO. OF RUNS = 15
```

```
THE EXPECTED NO. OF RUNS = 16.0000
```

```
15 OBSERVATIONS ABOVE K 15 BELOW
```

```
THE TEST IS SIGNIFICANT AT 0.7103
```

```
CANNOT REJECT AT ALPHA = 0.05
```

Som vi ser, kan vi ikke forkaste H_0 på 5%-nivået — tallene er trolig i tilfeldig rekkefølge.

7.9 Fortegnstester

Fortegnstest (sign test) er en ikke-parametrisk metode for å teste hypoteser når man arbeider med parvise sammenligninger. Testen benyttes ofte når man f.eks. skal vurdere om en medisinsk behandling har hatt virkning eller ikke. Man tester dette ved å sjekke om medianen til differansen mellom de to settene av observasjoner er lik 0 (eller for å si det på en annen måte: er "Etter" minus "Før" lik 0?).

Den generelle formen er

```
STEST [m] kolonne(r)
```


hvor m er den medianen du tester (hvis du ikke oppgir m , bruker Minitab 0). Alternativhypotesen (H_1) er vanligvis "medianen er ikke m " (tosidig test). For å lage en ensidig test brukes subkommandoen ALTErnative (se under TWOSample for bruk av denne subkommandoen).

Et eksempel: En del studenter er ikke enig i den karakteren de på eksamener eller fagoppgaver. De klager, og får ny evaluering. Resultatet av denne evalueringen blir stående. Resultater slik omsensur i noen forandring?

La oss ta utgangs punkt i følgende data¹⁷:

ROW	Sensur	Omsensur	Di ff.
1	2.7	2.6	-0.100000
2	3.5	3.9	0.400000
3	4.0	4.0	0.000000
4	1.9	1.6	-0.300000
5	2.7	3.6	0.900000
6	3.0	3.0	0.000000
7	3.0	2.9	-0.100000
8	3.7	3.0	-0.700000

Vi tester om medianen til differansen er lik 0:

```
MTB > STES t 'Di ff. '
```

```
SIGN TEST OF MEDIAN = 0.00000 VERSUS N. E. 0.00000
```

MEDIAN	N	BELOW	EQUAL	ABOVE	P-VALUE
Di ff. -0.05000	8	4	2	2	0.6875

Som synes, er signifikanssannsynligheten 0.6875 — hvis vi velger et signifikansnivå på 0.05, må dermed H_0 beholdes — det er ingen vits i å be om omsensur.

7.10 Mann-Whitney test

Mann-Whitney-testen brukes for å teste hypoteser og beregne konfidensintervall for medianer i to populasjoner. I Minitab er kommandoen for å utføre en Mann-Whitney-test

```
MANN-whitney [a] [k] kolonne1 kolonne2
```

hvor a kan være 1 (tester median1 > median2), 0 (tester median1 = median2) eller -1 (tester median1 < median2). Hvis du ikke skriver noen a , er 0 standard. k er konfidensnivået (95% hvis du ikke skriver noe). Kolonnene (må være to stykker) inneholder de to stikkprøvene — som ikke trenger å være sortert.

¹⁷ Disse dataene er fiktive - noe studenter som ikke er fornøyd med statistikkarakteren bør merke seg. . . .

Et eksempel:

En stikkprøve blant deltidsstudenter BI-Oslo 1982 ga dette resultatet for årslønn (hele tusen)¹⁸:

Menn:

82	90	100	102	103	106	108	110	110	120	125
130	130	130	135	140	140	144	150	150	152	160
160	165	165	170	175	176	180	185	216	300	

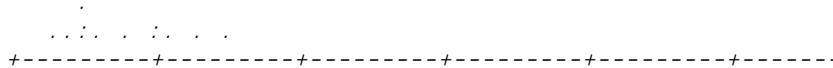
Kvinner:

91	96	99	100	100	105	110	120	120	124	130
140										

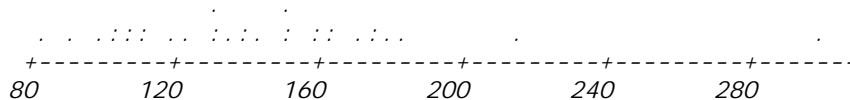
Først plotter vi tallene ved hjelp av DOTPlot-kommandoen (subkommandoen SAME gjør at aksene blir like):

```
MTB > DOTPlot C2 C1;  
SUBC> SAME.
```

Kvinner



Menn



Deretter kjører vi en Mann-Whitney test med nullhypotese median(menn) = median (kvinner), på 5% signifikansnivå.

```
MTB > MANN-Whitney C1 C2
```

Mann-Whitney Confidence Interval and Test

```
Menn      N = 32      MEDIAN =      140.00  
Kvinner   N = 12      MEDIAN =      107.50  
POINT ESTIMATE FOR ETA1-ETA2 IS      30.00  
95.0 PCT C. I. FOR ETA1-ETA2 IS ( 10.00, 49.99)  
W = 830.5  
TEST OF ETA1 = ETA2 VS. ETA1 N.E. ETA2 IS SIGNIFICANT AT 0.0037
```

Som vi ser, er testen signifikant på 5%-nivået (faktisk helt ned til 0.37% signifikansnivå — eller "med 99,83% sikkerhet"). Nullhypotesen forkastes: Mannlige deltidsstudenter tjener signifikant mer enn kvinnelige¹⁹.

¹⁸ Dette eksemplet er hentet fra Fred Wenstøp: *Statistikk og dataanalyse*, 1987, side 33.

¹⁹ Utskriften fra denne kommandoen skiller seg vesentlig fra den beskrevet i Wenstøp (1987), men hvis vi konsentrerer oss om *signifikanssannsynligheten* skulle resultatet bli det samme.

Utskrift

I dette (korte) kapittelet behandles endel kommandoer som er relatert til utskrift (både på papir og skjerm) av det som kommer opp på skjermen. Dette er nyttig i mange sammenhenger — f.eks. hvis du skal inkludere deler av utlistinger fra Minitab direkte i rapporter, fagoppgaver el.l.

7.11 Utskrift til printer

Hvis du gir kommandoen

```
PAPER
```

blir alt som kommer på skjermen også skrevet ut på skriver.

Denne kommandoen varierer litt mellom de ulike datamaskintypene; enkelte sender resultatet rett til skriveren, andre skriver til en fil som så skrives ut når den blir avsluttet (Se kapittel ?, ? og ?).

7.12 Utskrift til disk (fil)

På samme måte som du kan "dumpe" ting rett til en skriver, kan du i Minitab "dumpe" det som kommer opp på skjermen til en "Output file". Output files er filer som "tar vare på" det du gjør mens du sitter ved PC'en eller terminalen. Disse filene startes opp ved kommandoen

```
OUTFile 'filnavn'
```

(du *må* oppgi et filnavn) og avbrytes med kommandoen

```
NOOoutfile
```

Hvis du bruker det samme filnavnet flere ganger, blir output fra maskinen lagt til bunnen av den gamle filen. Dette betyr at en outfile kan bli ganske lang etterhvert — og derfor kan bruke mye tid og papir på skriveren. Derfor kan det være en tanke å "pusse litt" på disse filene ved hjelp av et editerings- eller tekstbehandlingsprogram før du sender dem til skriveren.

Detaljer om hva slags filnavn osv. varierer mellom de ulike datamaskintypene (Se kapittel ?, ?, ? og ?).

OUTFile-kommandoen har en subkommando: NOTErninal. Hvis du benytter den, blir output fra maskinen *kun* lagt til en fil — den kommer ikke opp på skjermen i det hele tatt. Dette kan være greit hvis du kjører store programmer — da er du ikke interessert i å sitte og se på at mengder av tall og bokstaver raser over skjermen (dessuten kan du få problemer med at Minitab stadig vekk stopper opp og spør "Continue?" når skjermen er full).

7.13 Continue?

Når du får listet opp noe på skjermen, stopper Minitab etter et visst antall linjer og spør "Continue?". Dette antallet linjer avhenger av hvilket type datamaskin du arbeider med (se egne kapitler). Du kan

regulere det med kommandoen OH (Output Height). Hvis du setter OH = 0, vil ikke skjermen stoppe i det hele tatt — og dette kan være greit hvis du allikevel tar kopi av alt som skjer på skjermen til en OUTFile.

Bredden på det som skrives ut på skjermen kan du regulere med kommandoen OW (Output Width). Dette gjelder særlig ting som PLOT o.l. — av og til kan det være greit å regulere bredden på dette, særlig hvis du skal "klippe" ut dette og ha det med i en rapport el.l. OW kan være mellom 30 og 132.

Det finnes også en kommando IW (Input Width), som regulerer bredden på de filene Minitab leser inn (data- og kommandofiler). Denne er vanligvis satt til 80, men kan være alt fra 9 opptil 160. Dette varierer fra datamaskin til datamaskin (se egne kapitler). IW-kommandoen er sjelden aktuell å bruke.

7.14 Utskrift fra enkelte kommandoer

BRIEF-kommandoen lar deg regulere hvor mye som skal skrives ut fra Minitabs side hvis du gir kommandoene REGress, DISCRiminant, ARIMA eller RLINe. Kun REGress og DISCRiminant er behandlet her — se HELP BRIEF for en beskrivelse av hvordan den oppfører seg sammen med de andre kommandoene. Den generelle formen på kommandoen er

BRIEF x

hvor x kan være 0, 1, 2, eller 3. BRIEF gjelder til du går ut av Minitab — eller til du gir en ny BRIEF-kommando. BRIEF er vanligvis satt til 2 (på noen datamaskiner til 1, se egne kapitler) når du starter Minitab.

For REGress-kommandoen finnes disse nivåene av BRIEF:

- 0 Ingenting skrives ut fra kommandoen — men det skrives ut til en OUTFile (forutsatt at du har gitt OUTFile-kommandoen på forhånd). Resultatet av alle subkommandoer skrives ut, så hvis du bare er ute etter en PREDICT kan denne opsjonen være grei å kjenne til.
- 1 Regresjonsligningen, koeffisientene, s , R^2 , justert R^2 og variansanalysedelen skrives ut.
- 2 I tillegg skrives en liste av "uvanlige" observasjoner ut (inklusive predikerte verdier og residualer)
- 3 I tillegg skrives alle observasjonene ut — inklusive predikerte verdier og residualer.

For DISCRiminant-kommandoen finnes disse nivåene av BRIEF:

- 0 Tilsvarende BRIEF under REGress.
- 1 Kun klassifikasjonsmatrisen skrives ut
- 2 Standard utskrift—se pkt. ?, side 95 for et eksempel.
- 3 I tillegg til det standard-utskriften, skrives også gjennomsnitt, standardavvik og covarians-matrisen ut, både for hver enkelt kategori og for hele populasjonen (pooled).
- 4 I tillegg skrives alle observasjonene ut med klassifikasjon og klassifikasjonssannsynlighet.

Minitab er i utgangspunktet interaktivt — det vil si at du gir kommandoer til programmet, som så gir deg "svar" for hver kommando. Det motsatte av "interaktivt" er "batch-orientert" (satsvis) — det vil si at du må skrive programmer (f.eks. kjeder av kommandoer) som så kjøres igjennom av statistikkprogrammet²⁰.

Interaktiv bruk av Minitab er greit nok — men det kan ha sine ulemper. Hvis du bruker Minitab endel, vil du fort oppdage at du ofte gir de samme kommandoene gjentatte ganger — og at du ofte gir den samme sekvensen av kommandoer gjentatte ganger. Et eksempel kan være at du kjører den samme analysen av mange kolonner — eller du kan f.eks. ha flere datafiler som du kjører de samme analysene på.

Minitab har etterhvert fått ganske gode muligheter til å lage små "programmer" som består av grupper av kommandoer. Disse programmene (kommandofilene, makroene) kan du lagre — og starte dem ved behov. Hvis du går inn for å lære deg denne delen av Minitab, kan du spare deg selv for mye kjedelig arbeid — for ikke å snakke om hvor mye raskere det går. . . .

Noen eksempler:

- det før nevnte eksemplet med "faste" analyser — ting du gjør på samme måte med en rekke datasett eller kolonner. Ved å bruke kommandofiler kan du til en viss grad lage deg "dine egne" Minitab-kommandoer.
- "prøvekjøring" av komplekse analyser på små datamengder. Du tar ut en liten del av datamaterialet du skal analysere, og finner ut hvilke kommandoer du skal kjøre, og i hvilken rekkefølge. Kommandoene lagrer du med JOURNAL-kommandoen — og etter litt editering kan du kjøre programmet på hele datamengden. Fordelen med dette er at du kan gjøre noe annet mens programmet går — du slipper å sitte og vente på svar fra maskinen på hver enkelt kommando.
- overføring av data. Hvis du skal flytte data fra en datamaskin til en annen kan du ikke benytte lagring ved hjelp av SAVE²¹. Du må bruke WRITE — og dermed mister du slike ting som kolonnenavn osv. Hvis du lagrer dataene som en datafil og samtidig lager en liten kommandofil som navngir kolonnene, kan du overføre begge deler til den andre datamaskinen, kjøre kommandofilen — og du har worksheet'et som det var før du lagret det.

7.15 Et raskt lite program.

La oss ta et raskt lite eksempel. Du har et datasett med 30 kolonner. Du ønsker å tabulere den første

²⁰Her gjennomgår jeg kun Minitab's enkleste måte å programmere på. Minitab versjon 9, som i skrivende stund (November 1993) kun finnes for Digital's VAX-maskiner og for Microsoft Windows, har et mer avansert programmeringssystem, kalt %Macros. Med dette systemet kan man programmere "skikkelig", med DO..WHILE løkker, IF..THEN instruksjoner og andre ting programmerere liker.

²¹Med et lite unntak; f.o.m. utgave 6.1.1 av Minitab kan worksheets lagres med format PORTABLE - som muliggjør overføring til andre maskiner.

kolonnen mot hver av de andre. Hvis du skal gjøre dette manuelt, må du gi 30 kommandoer — og vente på hver enkelt av dem. Ved hjelp av kommandofiler kan du gjøre det f.eks. slik:

```
MTB > LET K1 = 2      (Setter K1 til 2)
MTB > STORE           (Lagre de følgende linjer)
STOR> TABLE C1 CK1  (Tabuler C1 mot CK1, dvs C2)
STOR> LET K1 = K1 + 1 (Øk K1 med en)
STOR> END            (Slutt på lagring av program)
MTB > EXECute 29     (Kjør programmet 29 ganger)
```

Etter denne sekvensen vil Minitab tabulere C1 mot alle de andre kolonnene, en av gangen. (Du bør imidlertid ha brukt OUTfile-kommandoen med subkommando NOTEriginal før du starter programmet ditt — ellers vil alle tabuleringene rase over skjermen fortere enn du klarer å lese dem).

Legg merke til hvordan vi bruker K1 (eller en hvilken som helst konstant) i eksemplet. K1 betyr 2,3,4,5. . . osv. etterhvert som programmet blir kjørt, fordi den siste linjen i programmet øker K1 med 1.

Nå over til mer detaljert forklaring av hvordan du lager programmer i Minitab.

7.16 Hvordan lage kommandofiler

Du kan lage kommandofiler til Minitab på 3 måter:

- ved å bruke STORE-kommandoen
- ved å bruke JOURNAL-kommandoen
- ved å skrive (eller forandre på) programmet ved hjelp av en editor eller et tekst-behandlingsprogram

I alle tilfellene lages programmet som en vanlig, lesbar fil (gjerne kalt ASCII-fil på PC eller Macintosh).

7.16.1 STORE-kommandoen

STORE-kommandoen er demonstrert i eksemplet i begynnelsen av dette programmet. Den generelle formen på denne kommandoen er

```
STORE ['filnavn']
```

Hvis du ikke oppgir noe filnavn, vil Minitab benytte et standard filnavn. Hva dette er varierer mellom de ulike datamaskintypene (se kapittel ?, ?, ? og ?). Hvis du skal oppgi noe filnavn, må du sette det innenfor apostrofer, på samme måte som f.eks. SAVE- eller READ-kommandoen.

Når du har gitt STORE-kommandoen, forandrer Minitab's prompt (klartegn) seg fra "MTB >" til "STOR>". Alle kommandoer du deretter gir blir ikke utført, bare lagret til fil. Du avslutter STORE-kommandoen med END (på samme måte som f.eks. SET-kommandoen). Etter at du er ferdig med STORE-kommandoen kan du kjøre kommandofilen med EXECute-kommandoen.

7.16.2 JOURNAL-kommandoen

JOURNAL-kommandoen minner om STORE-kommandoen. Den eneste forskjellen er at JOURNAL utfører kommandoene i tillegg til å lagre dem. Generell form på JOURNAL-kommandoen er

```
JOURnal ['filnavn']
```

Hvis du ikke oppgir noe filnavn, vil Minitab benytte et standard filnavn. Hva dette er varierer mellom de ulike datamaskintypene (se egne kapitler). Hvis du skal oppgi noe filnavn, må du sette det innenfor apostrofer, på samme måte som f.eks. SAVE- eller READ-kommandoen.

Når du har gitt JOURNAL-kommandoen, lagrer som sagt Minitab alle kommandoer du gir fra tastaturet inn i en fil. Dette gjøres helt til du slår av funksjonen ved kommandoen NOJOURNAL. På denne måten fungerer JOURNAL som en "båndspiller" som tar opp alt du gjør. Dette kan være svært greit hvis du f.eks. skal kjøre de samme kommandoene på flere datasett — først gjør du det du skal gjøre med JOURNAL på — deretter avslutter du med NOJOURNAL — og kan kjøre kommandofilen JOURNAL har laget ved EXECute-kommandoen.

NB! JOURNAL-kommandoen lagrer filer med litt anderledes filtype enn STORE-kommandoen (se egne kapitler). Det kan derfor hende at du må gi filen(e) nytt navn/type før du kan kjøre dem — eller du må inkludere filtypen i EXECute-kommandoen.

7.16.3 Bruk av editorer

En editor er et "program for å skrive programmer" — en slags tekstbehandlingsprogram beregnet på programtekster. Det finnes forskjellige editorer for forskjellige datamaskiner. Ofte kan du benytte tekstbehandlingsprogrammer som editorer — som regel må du da lagre programmet på en spesiell måte (f.eks. som "DOS-tekst" i WordPerfect på en IBM-kompatibel PC). Kjente editorer er f.eks. Brief, EMACS, Qedit eller Kedit på IBM-kompatible PC'er, Xedit på IBM stormaskin og VI eller EMACS på VAX. På en Macintosh kan du bruke MacWrite eller et annet tekstbehandlingsprogram og lagre filen som "Text Only".

Kommandofiler i Minitab er akkurat som datafilene; vanlige lesbare filer i datamaskinens "standardformat". Dette betyr at du kan benytte en editor for å lage eller forandre dem. Programmene ser ut som lister av kommandoer (uten "MTB >" o.l. foran hver kommando). Programmeringseksemplet i begynnelsen av dette kapitlet ville sett slik ut hvis vi hadde listet opp filen på skjermen:

```
TABLE C1 CK1  
LET K1 = K1 + 1  
END
```

7.17 Hvordan kjøre kommandofiler

For å starte kommandofiler inne fra Minitab bruker du kommandoen EXECute. Den generelle formen for denne kommandoen er

```
EXECute ['filnavn'] [n]
```

hvor n er antall ganger kommandofilen skal kjøres (standardverdi er 1 gang). Hvis du ikke setter noe filnavn, vil Minitab lete etter det samme filnavnet som blir laget hvis du gir STORE-kommandoen uten filnavn (se kapitlene om de ulike datamaskintypene).

Fra og med versjon 7.1 kan du lage en Minitab kommandofil med et spesielt navn (på IBM PC: STARTUP.MTB), som kjøres automatisk hver gang du starter opp Minitab.

7.18 Noen tips.

Her er noen tips for "programmere". De er meget kortfattede, fordi jeg regner med at det kun er et fåtall av denne manualens lesere som har interesse av dem. Men i alle fall — lykke til!

- BATCH øverst og TSHARE nederst i programmet (kommandofilen) gjør at du slipper "Continue?"-meldinger o.l. BATCH gjør at teksten vil rase over skjermen — noe du som regel ønsker i et program. TSHARE (time share) slår på igjen "Continue"-funksjonen.
- OUTFILE-kommandoen er grei å bruke — fordi resultatet av en kommandofil med mengder av kommandoer ikke så lett lar seg lese på skjermen "i fart". Bruk også subkommandoen NOTERminal — så slipper du å se noe på skjermen i det hele tatt (og får ikke trøbbel med at Minitab stopper etter et visst antall linjer og spør "Continue?").
- hvis du bruker OUTFile med NOTERminal, kan det bli problematisk hvis du ikke avslutter kommandofilen med NOOUfile.
- du kan i tillegg gi kommandoen NOECho på begynnelsen av en kommandofil — da vil heller ikke kommandoene vises på skjermen etterhvert som de blir eksekvert. Husk imidlertid å bruke ECHO på slutten. Du kan også benytte ECHO med en tekst på enkelte steder i kommandofilen — så har du en mulighet for å følge med hvordan det går (dette er særlig nyttig for å finne feil eller sjekke at kommandofiler ikke går i en "uendelig løkke").
- NOTE-kommandoen er grei å benytte i store programmer — gir deg en mulighet til å beholde oversikten
- du kan også gi kommentarer *etter* hver kommando ved å sette "#" (IBM stormaskin og VAX kan benytte andre tegn, se kapittel ? og ?) før kommentaren. Et eksempel:

```
INFO      # Dette er en kommentar
```

- hvis du skriver programmer som behandler store mengder data, kan det ofte være en fordel å slette kolonner med data du ikke har bruk for lenger. Dette kan få Minitab til å jobbe raskere — og det kan hindre at du "stanger i taket" på antall dataelementer/kolonner.
- hvis du setter subkommandoene et par anslag inn på linjen blir programmene lettere å lese. Sammenlign

```
TABLE C6 C7;          TABLE C6 C7;
ROWPercents;          ROWPercents;
COLPercents.          COLPercents.
DESCRibe C8;          DESCRibe C8;
```


- det er mulig å skrive kommandofiler som kjører andre kommandofiler. Dette gjøres ved å ha EXECute-kommandoen inne i kommandofilen. Det er imidlertid to ting du må passe på her
 - at du ikke lager for mange "kommandofilnivåer" (filer som kaller filer som kaller filer etc.). Hvor mange "lag" (levels) av kommandofiler du kan ha varierer mellom de ulike datamaskintypene (se kapittel ?, ?, ? og ?), men som regel er det 5 eller 6, inklusive "toppfilen" og eventuell OUTFile.
 - at du ikke lager "endeløse løkker" (endless loops²²) Dette kan skje hvis du har en kommandofil 1 som kaller kommandofil 2, hvorpå kommandofil 2 kaller kommandofil 1 igjen. . . . Hvert av disse kallene lager et nytt "lag" av kommandofiler — og du kan få noen merkelige feilmeldinger. . . .
- fra og med versjon 7.1 av Minitab finnes kommandoen YESNo, som lar deg stoppe en makro slik at brukeren kan svare YES eller NO på et spørsmål. Dette kan f.eks. foregå slik:

```

NOECHO
ECHO Ønsker du en oversikt over dataene (Yes/No)?
YESNo K1
EXECute 'program2' K1

```

Verdien av K1 (eller hvilken som helst K) blir 1 hvis svaret er et ord som begynner på Y eller y, 0 hvis svaret begynner på N eller n. Hvis svaret er noe annet, kommer det en tekst "*Please answer Yes or No*", og prosessen gjentas. Dette gjentar seg inntil 5 ganger — da regnes svaret for å være 0. Deretter kan du kjøre et annet program K1 ganger — og hvis K1 er 0, vil jo ikke dette programmet kjøres i det hele tatt.

- hvis du ønsker å la brukeren av programmet lese et tall inn (f.eks. kan du jo gi beskjed om hvor mange ganger en programfil skal kjøres) kan du benytte SET-kommandoen med subkommandoen NOBS. Et eksempel på hvordan dette kan gjøres er vist i programmeringseksemplet i neste punkt.

7.19 Programmeringseksempel: enarmet banditt

Som et eksempel på bruk av programmering — la oss fortsette på simuleringseksemplet fra punkt ?. Vi fant der at for en enkelt simulering fikk vi gevinst. Vi regnet også ut at forventningen av å spille på den enarmede banditten var negativ. Ved hjelp av to relativt små programmer kan vi simulere dette flere ganger på en effektiv måte.

Vi begynner med dette programmet, som vi kan kalle STARTSIM:

```

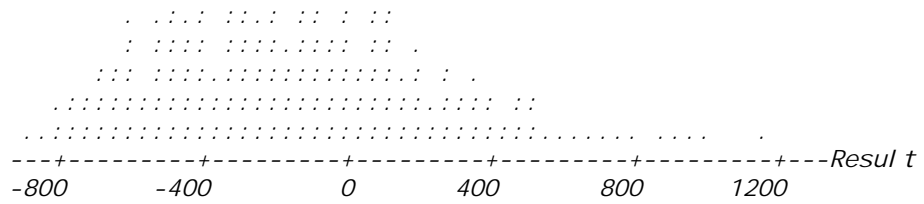
NOECHO                                # ikke skriv kommandoene ut på skjermen
BATCH                                  # setter Mini tab i BATCH modus

```

²²Definisjonen på "endless loop" (iflg. Ashton Tate's dBase III-manual):

Endless loop: see "Loop, Endless".

Loop, Endless: see "Endless loop".



MTB > DESCRibe 'Result'

	<i>N</i>	<i>MEAN</i>	<i>MEDIAN</i>	<i>TRMEAN</i>	<i>STDEV</i>	<i>SEMEAN</i>
<i>Result</i>	1000	-154.6	-185.5	-166.3	358.4	11.3
	<i>MIN</i>	<i>MAX</i>	<i>Q1</i>	<i>Q3</i>		
<i>Result</i>	-863.0	1160.0	-430.5	88.0		

Som vi ser, nærmer dette (i tråd med sentralgrenseteoremet) seg en normalfordeling med negativ forventning. De mer spillegale har antagelig forlenget lagt merke til at den høyeste gevinsten av disse 1000 forsøkene spillet var 1160 — noe som kanskje kan gi en viss innsikt i hvorfor så mange mennesker synes spilleautomatene på danskebåten og i Las Vegas er så fascinerende. De mer forsiktige ser at i den mest uheldige omgangen satt man faktisk igjen med kun 137 kroner av den tusenlappen som gikk med i innsats. . . .

Minitab og andre programmer

I dette kapittelet tar vi for oss hvordan Minitab fungerer sammen med andre programmer det kan være naturlig å bruke — f.eks tekstbehandling eller regneark.

7.20 Data fra/til regneark

Minitab på IBM PC har gode muligheter for å hente data fra et Lotus-kompatibelt regneark. "Lotus-kompatibelt" vil si at regnearket kan lagre filer i Lotus 1-2-3 eller Lotus Symphony format. Regneark som Excel (både PC og Macintoshversjonen), Quattro, SuperCalc4 og SuperCalc5 kan gjøre dette.

Konverteringen mellom Minitab og Lotus-format kan gjøres begge veier (til og fra Minitab/regneark). Fremgangsmåten er litt forskjellig mellom Minitab versjon 6 og 7.

I *Minitab versjon 7* (og senere versjoner) gjøres dette med subkommandoen LOTUs til kommandoene SAVE og RETRIEve²⁴. Et eksempel:

```
MTB > SAVE 'mi nedata' ;  
SUBC> LOTUs.
```

Dataene vil lagres som en fil med endelsen .WK1, som er det formatet som benyttes i Lotus 123 versjon 2. Du kan endre dette til .WKS-format (Lotus 123 versjon 1) eller Lotus Symphony-formatene .WRK eller .WR1 i MSETUP-programmet.

7.21 FORTRAN-formaterte data

I riktig gamle dager, dvs. sånn omtrent 1970, brukte man hullkort til å lagre data på. Et hullkort hadde plass til 80 tegn — det er bl.a. derfor det fremdeles er 80 tegn i bredden på en dataskjerm. Hullkortet har gitt opphav til FORTRAN-formaterte data, som er en måte å spesifisere hvordan data ser ut. Med visse "dialektforskjeller" blir dette formatet brukt i svært mange statistikkpakker, f.eks. i SPSSX (Statistical Package for the Social Sciences, Extended). Det er også et meget brukt format av offentlige datakilder — f.eks. Statistisk Sentralbyrå.

Sett at vi har en fil som inneholder disse tallene:

```
217670. 038. 520100  
117263. 041. 021 0  
117072. 041. 020 0  
117469. 040. 020 0  
216652. 038. 020 0  
119487. 045. 021 0  
117967. 042. 020 0  
117369. 042. 022 80  
118070. 042. 021 5
```

²⁴I Minitab versjon 6 for PC er det et eget program ved navn MTB123.EXE som konverterer Minitab worksheets til og fra Lotus-filer.

Hvis du ser litt nøyere etter, vil du kjenne igjen de 9 første radene fra "studentdataene" — se liste bakerst i manualen. Det første sifferet på hver linje er kolonnen 'Kjonn' — 1 for menn, 2 for kvinner. De tre neste sifrene er høyden i centimeter, deretter følger 4 sifre (inklusive den plassen desimalpunktumet tar) som er kroppsvekten. Deretter nok en "kolonne" med 4 sifre (inklusive desimalpunktumet). Så alder (2 sifre), til slutt antall sigaretter pr. uke (3 sifre).

For å lese inn dette i Minitab kan vi gi disse kommandoene (forutsetter at dataene ligger i filen FDATA.DAT):

```
MTB > READ 'fdata' c1-c6;
SUBC> FORMat (f1. 0, f3. 0, 2f4. 0, f2. 0, f3. 0).
          9 ROWS READ
```

ROW	C1	C2	C3	C4	C5	C6
1	2	176	70	38. 5	20	100
2	1	172	63	41. 0	21	0
3	1	170	72	41. 0	20	0
4	1	174	69	40. 0	20	0
.

```
MTB > _
```

Se litt nærmere på FORMat-subkommandoen; den er langt fra så kryptisk som den kan se ut ved første øyekast. F'en betyr "field" — eller felt. Tallet etter f'en angir hvor mange tegn feltet (kolonnen, variabelen) opptar. Hvis det står et tall rett foran f'en, indikerer den hvor mange like felter som følger rett etter hverandre — 2f4.0 betyr altså to felter, hver med 4 tegn. Du kan også hoppe over tegn — dette gjøres med et tall (antall tegn som skal "hoppes over") og bokstaven "X".

FORMat-subkommandoen kan benyttes med kommandoene READ, WRITe, SET, INSErt og PRINt.

På samme måte som med det vanlige Minitab dataformatet (dvs. med mellomrom eller komma mellom hver kolonne) er det én observasjon (respondent) pr. linje. Ved det vanlige Minitab-formatet benyttes "&" som fortsettelsestegn hvis linjen er for lang — hvis dataene er FORTRAN-formattert benyttes skråstrek ("/"). Hvis datafilen hadde sett slik ut for hver respondent:

```
217670. 038. 5
20100
```

ville kommandoen altså sett slik ut:

```
MTB > READ 'fdata' c1-c6;
SUBC> FORMat (f1. 0, f3. 0, 2f4. 0/f2. 0, f3. 0).
```

Alfanumeriske data (dvs. bokstaver og tall) kan også leses inn ved hjelp av FORMat-subkommandoen. Hvis du har disse dataene liggende i filen A_DATA:

```
PER 110
PAAL 34
ESPEN 76
```

kunne denne kommandoen brukes

```
MTB > READ 'adata' c1 c2;  
SUBC> FORMat (a6, f3.0).
```

A'en her står for "alfanumerisk", og tallet 6 betyr at de 6 første tegnene fra den begynnelsen av hver linje skal inn i kolonne 6.

7.22 Minitab og tekstbehandlingsprogrammer

Det er ofte aktuelt å benytte Minitab sammen med et eller annet tekstbehandlingsprogram — enten for bruke tekstbehandlingen til å redigere Minitab kommando- eller datafiler, eller for å "ta inn" Minitab-utskrifter (ved hjelp av OUTFile-kommandoen) i tekster. Her er noen tips om bruk av Minitab sammen med tekstbehandlingsprogrammer, delvis som et resultat av arbeidet med å skrive denne manualen.

- hvis du skal skrive en data- eller kommandofil i et tekstbehandlingsprogram for senere å benytte den i Minitab; vær sikker på at du lagrer teksten i et format som Minitab kan lese. Alle tekstbehandlingsprogrammer har en kommando eller opsjon som tillater deg dette. På IBM PC'er kalles dette gjerne "DOS-format", "ASCII-text" eller "Print to File". På Macintosh kalles det gjerne "ASCII-text" eller "Text only".
- hvis du ønsker å ta tekst fra Minitab inn i et tekstbehandlingsprogram, må teksten lagres med OUTFile-kommandoen. Hvis du skal hente inn dataene kan det være fristende å benytte WRITE-kommandoen — men hvis det er snakk om mange rader kan det lønne seg å lage en OUTFile, for deretter å benytte PRINT-kommandoen.
- hvis du henter inn OUTFiles til et tekstbehandlingsprogram: vær oppmerksom på at alle mellomrom i OUTFile'n (også de som ser ut som innrykk og tabulator tegn) er vanlige mellomrom. Dette betyr at du kan få problemer med tabeller o.l. som "logrer" hvis du benytter proporsjonalskrift (Times etc.). Løsning: enten å erstatte mellomrom med tabulator tegn (tungvint og upresist) eller å gå over til en "monospaced" skrifttype på den teksten som består av tabeller o.l. (slik det er gjort i denne manualen).
- Minitab's høyoppløselige grafikk kan som regel kun med vanskelighet hentes inn i tekst. Det finnes endel programmer som tillater deg å "fotografere" skjermen til en fil (Word Perfect's GRAB er et av dem, ellers finnes det endel gratisprogrammer). Som regel er det lettere (og mer fleksibelt) å lagre dataene fra Minitab inn i et regneark eller grafikkprogram og bruke grafikken derfra.

7.22.1 Eksempel: Tekst fra Minitab til Wordperfect 5.x

La oss ta et lite eksempel: Vi skal foreta endel kommandoer i Minitab, som vi ønsker å ta med inn i et dokument skrevet i WordPerfect (versjon 4.2 eller 5.x) på en IBM-kompatibel PC. Jeg forutsetter at det finnes en harddisk i denne PC'en og at harddisken har betegnelsen "C:".

Fremgangsmåten kan være som følger:

- Start opp Minitab. Gi kommandoen

OUTFile 'mintekst'

I stedet for "mintekst" setter du naturligvis et annet filnavn hvis du ønsker det.

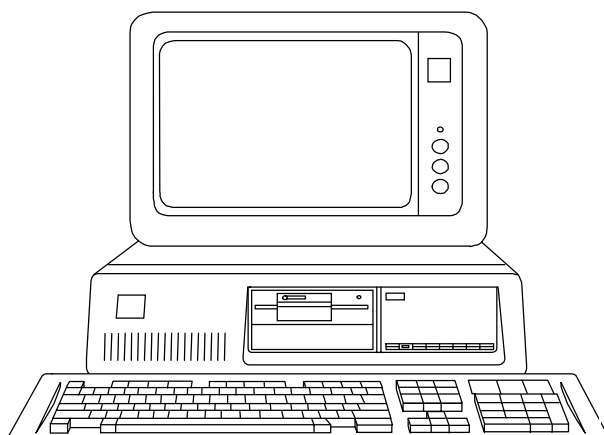
- Gjør det du skal i Minitab. Alle kommandoer du gjør — og resultatet av dem — blir lagret i en fil med navn MINTEKST.LIS. Når du er ferdig med det du skal gjøre, gi kommandoene

*NOOUTfile
STOP*

- Når du er ute av Minitab (tilbake til "C:\>" e.l.) starter du opp WordPerfect og henter inn det dokumentet du holder på med. Du flytter markøren i WordPerfect til den står på den plassen i dokumentet du vil ha inn Minitab-teksten. Så trykker du Control-F5 (dvs. holder ned Control-tasten og trykker funksjonstast 5). Du får opp en meny hvor alternativ 1 er "DOS-tekst". Velg dette alternativet (ved å trykke 1-tallet).
- I den menyen som kommer opp, velg alternativ 3 (Hent dokument, CR/LF blir Fast linjeskift).
- Kommandoene fra Minitab-sekvensen er nå lest inn i dokumentet ditt. Gå igjennom dem og slett ting som ikke er nødvendige (f.eks. vil du sannsynligvis ha en masse "faste sideskift" i teksten).
- Hvis du benytter proporsjonalskrift (dvs. en skrifttype på din skriver hvor ikke alle bokstavene er like brede): Skift til en annen skrifttype for de delene av teksten som er direkte utskrift av Minitab-kommandoer. Ellers vil du som tidligere nevnt få endel problemer med at tekst og tall i kolonner ikke blir stående rett under hverandre — men i stedet "slynger seg" frem og tilbake.

Minitab på IBM PC'er

zzzz Den nyeste versjonen av Minitab for IBM-kompatible PC'er er versjon 7.2. Versjon 7.2 er den samme som versjon 7.1 med noen små rettelser. Du kan imidlertid godt bruke versjon 6.1 (den største forskjellen er at dataeditoren ikke er der), eller versjon 5.1.1 eller 5.1.3 (disse har færre kolonner og ingen muligheter for høyoppløselig grafikk). Versjonen før 5.1.1 het 1982.1 — den kan fremdeles brukes, men såpass mange kommandoer er forskjellige fra de man finner i senere versjoner at det nok er lurest å oppdatere. . . .



Minitab versjon 6.1.1 krever at du har minst 450 Kb ledig internhukommelse (dette kan du sjekke med CHKDSK-kommandoen)²⁵. Harddisk anbefales på det varmeste — men det er mulig å kjøre Minitab fra diskett også (se eget punkt).

Minitab versjon 7.1 og 7.2 krever *minst* 480 Kb ledig internhukommelse (dvs. at man må ha minst 640 Kb totalt) og harddisk (programmet tar ca. 3.2 Mb harddiskplass, kan reduseres til ca 2.8 Mb hvis du tar vekk endel filer du ikke trenger). Med andre ord — kjøp deg en harddisk (det vil du trenge i alle fall — de fleste "skikkelige" programmer idag trenger en harddisk).

Minitab kan kjøres på XT-, AT-, 386- og PS/2-maskiner. Hvis du kjører store datamengder og har en "gammeldags" XT, går det ikke fort (du får tid til å se at fingerneglene dine vokser i mellomtiden). Matematisk koproessor (såkalt 80x87-prosessor) er en fordel, men ingen nødvendighet.

Du starter Minitab ved å stå i den underkatalogen (subdirectory) som inneholder Minitab-programmet og skrive

MI NI TAB

og trykke Enter (returtasten).

Hvis du får problemer med at Minitab tar for mye internhukommelse, kan du forsøke å starte Minitab slik:

MI NI TAB /N

Dette fjerner muligheten for høyoppløselig grafikk — men frigjør noe internhukommelse. Hvis ikke dette hjelper, kan du starte Minitab med denne kommandoen:

MI NI TAB /N /E n

hvor *n* er et tall mellom 1000 og 32000 (jo større tall du skriver, jo mer internhukommelse frigjøres). Imidlertid reduseres antall mulige dataelementer med $n/4$. Dette vil si at hvis du starter Minitab med $n=32000$, kan du bare ha 8000 dataelementer i ditt tallmateriale, i motsetning til ca. 16000 som du kan

²⁵ Dette betyr bl.a. at hvis du bare har 512 Kb internhukommelse, kan du ikke kjøre senere versjoner av DOS enn 3.3 (fordi DOS 4.0 tar for mye internhukommelse).

ha ellers.

7.23 Fullskjerm dataeditor

Minitab for IBM-kompatible PC'er har fra og med versjon 7.1 en fullskjerm dataeditor. Dette er et regnearklignende verktøy som lar deg manipulere direkte med kolonner og rader.

Dataeditoren kommer du inn i ved å trykke ESC (Escape-tasten) når du er inne i Minitab. Den minner om et regneark — med unntak av at den kun har tall og bokstaver (ikke formler) i cellene. Jeg vil ikke beskrive den i detalj her — den er allikevel såpass lett å benytte at de fleste med minimal regnearkerfaring vil kunne ta den i bruk direkte. Følgende momenter bør du merke seg:

- editoren virker direkte på det worksheet (dvs. de data) du til enhver tid har i arbeidsområdet. Imidlertid blir ikke endringer lagret før du går tilbake til Minitab's vanlige skjermbilde og bruker SAVE-kommandoen.
- hvis du trykker F1 (funksjonstast 1) når du er inne i dataeditoren, får du opp en meget informativ hjelpefunksjon.
- F10 (funksjonstast 10) gir deg en meny med endel funksjoner

På enkelte PC'er (særlig eldre XT'er) vil denne editoren kunne oppleves som nokså treg (det tar tid før maskøren flytter seg når du trykker piltaster osv.). Dette skyldes at den er programmert til å gi signaler til skjermen gjennom noe som kalles BIOS — istedet for å skrive rett til skjermen. Fordelen med dette er at den vil fungerer for alle typer dataskjermer — men ulempen er altså at det går tregere.

7.24 Filnavn og -typer (IBM PC)

Minitab sørger selv for å lage riktig filtype for de spesielle filene som brukes. Disse filtypene er

- .MTW** for Minitab worksheets (lagres med SAVE-kommandoen, hentes med RETRIEve-kommandoen). Hvis du ikke spesifiserer noe filnavn i det hele tatt, bruker Minitab filnavnet MINITAB.MTW.
- .MTP** for worksheets som er lagret med subkommando **PORTable**. Disse kan overføres til andre datamaskiner og hentes inn med RETRIEve-kommandoen der. (Kun fra og med Minitab versjon 7.1).
- .DAT** for redigerbare datafiler (datafiler som hentes og lagres med henholdsvis READ- og WRITe-kommandoen).
- .LIS** for OUTFile-filer. OUTFile-kommandoen krever et filnavn — så noe "standard" filnavn finnes ikke.
- .MTB** for kommandofiler. Dette gjelder kommandofiler som lagres ved hjelp av STORe-kommandoen, og kjøres ved hjelp av EXECute-kommandoen. Hvis du ikke bruker noe filnavn (typisk ved små kommandoløkker) bruker Minitab filbetegnelsen MINITAB.MTB.
- .MTJ** for JOURnal-filer (filer laget med JOURnal-kommandoen). Hvis du ikke

spesifiserer noe filnavn, bruker Minitab filnavnet MINITAB.MTJ. **NB!** EXECute-kommandoen ser etter filer med type .MTB hvis du ikke spesifiserer noe annet — du må derfor bruke DOS' RENAME-kommando hvis du ønsker å slippe å spesifisere filtypen.

Det anbefales å bruke disse filtypene for de ulike Minitab-filene. Det er imidlertid mulig å bruke andre filtyper, men da må både filnavn og filtype spesifiseres i den aktuelle Minitab-kommando.

7.25 Kapasitetsbegrensninger (IBM PC)

I versjon 7 og 6 av Minitab for IBM PC'er og kompatible finner vi de følgende kapasitetsbegrensninger:

Størrelse på worksheet:	16,000 celler (endel mindre i praksis, fordi Minitab trenger endel ledige celler til mellomlagring ved beregninger, sortering og annet).
Max. antall kolonner:	100 (versjon 5: 50 kolonner)
Max. antall konstanter:	100
Max. antall matriser:	15
OW (Output Width):	79 tegn i bredden (kan maksimalt settes til 132)
OH (Output Height):	24 linjer
IW (Input Width):	160 tegn (maks. bredde på datafiler).
BRIEF:	2 (utskrift av resultat av REGress og endel andre kommandoer)
HEIGHT:	17 linjer (høyden på PLOT)
WIDTH:	57 tegn (bredden på PLOT)
Maks. antall åpne filer:	6 (dette vil si at du kan ha 6 kommandofil-lag (5 hvis du har en OUTFILE åpen)).

7.26 SYSTEM-kommandoen

SYSTEM-kommandoen lar deg "gå ut i" DOS (eller, rettere sagt, et subset av DOS) uten at du avslutter Minitab-kjøringen. For å komme tilbake til Minitab igjen, skriver du EXIT.

Det anbefales å bruke SAVE-kommandoen før du bruker SYSTEM. Dette fordi det er lettere å "crashe" datamaskinen (f.eks. ved at du starter et eller annet program som er for stort til at maskinen klarer å kjøre både det og Minitab samtidig). Da er det jo trist om du ikke har lagret. . . .

Her er et eksempel:

```
MTB > SYSTEM
To return to Mini tab, type "EXIT" at the DOS prompt
C: \M I N I T A B >
.
.
(DOS-kommandoer kjøres. . . .)
```

```
C:\MINITAB> EXIT
```

```
MTB > _
```

(du er tilbake i Minitab igjen)

I Minitab versjon 7.1 og senere kan du bruke DOS-kommandoene

<i>DIR</i>	(lister opp filnavn m.m.)
<i>TYPE</i>	(lister opp innholdet i en enkelt fil)
<i>CD</i> eller <i>CHDIR</i>	(forandrer directory (underkatalog) på harddisk/diskett)

direkte inne fra Minitab uten å gå om SYSTEM-kommandoen.

7.27 Grafikk på IBM PC

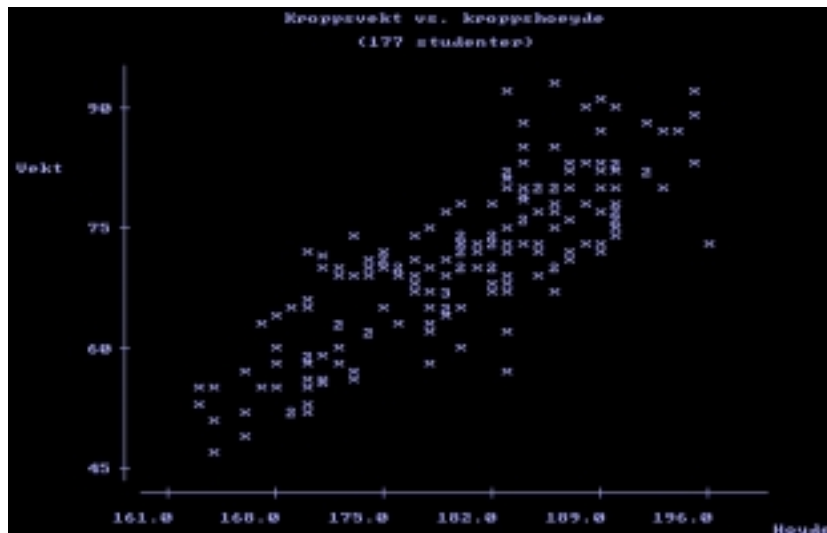
F.o.m. versjon 6.1.1 av Minitab er det mulig å kjøre høyoppløselig grafikk ("skikkelig" grafikk) på Minitab. Dette gjøres ved de samme kommandoene som du vanligvis benytter til grafikk, med den forskjell at du setter en G foran:

GPLOT
GHISTogram
GBOXplot

Du kan forandre på grafikken ved kommandoen GOPTIONS — uten at jeg skal gå nærmere inn på den her.

For at du skal kunne kjøre grafikk, må du ha en grafisk skjerm (CGA, Hercules, EGA, MCGA, VGA etc.). Minitab må vite hva slags skjerm du har — og det forteller du ved å kjøre programmet MSETUP. Her forteller du også hva slags skriver du eventuelt har, slik at du kan skrive ut grafikken.

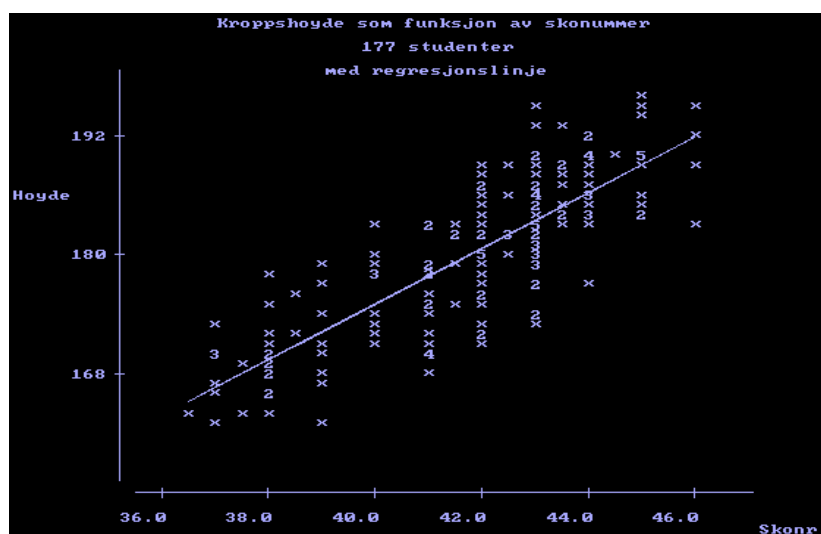
Grafikken skriver du ut (krever at du har en grafisk skriver knyttet direkte til PC'en) ved å få opp grafikken på skjermen, og så trykke "P" (dvs. bokstaven P).



GPLOT-kommandoen har også en meget hendig subkommando; LINE. Med LINE kan du få tegnet opp en linje mellom tallpar i kolonner. Hvis vi tar utgangspunkt i eksemplet med regresjon mellom høyde og skonummer — hvor vi la de prediserte verdiene inn i kolonne 29 — kan du gi disse kommandoene:

```
MTB > GPL0t 'Hoyde' 'Skonr.' ;
SUBC> TITLE 'Kroppshoyde som funksjon av skonummer' ;
SUBC> TITLE '177 studenter' ;
SUBC> TITLE 'med regresjonslinje' ;
SUBC> LINE C29 'Skonr.' .
```

og få ut denne grafen:



7.28 Hvis du ikke har harddisk. . . .

Harddisk er en stor fordel på en PC — og det gjelder de fleste programmer. Imidlertid er det **mulig** å kjøre Minitab (i hvert fall tidligere versjoner enn 7.1) fra en diskettstasjon også — men da vil du stadig få spørsmål av denne typen:

```
* NOTE * Please insert MINITAB DISK 1.
Press <ENTER> to continue or '.' to quit Minitab ==>
```

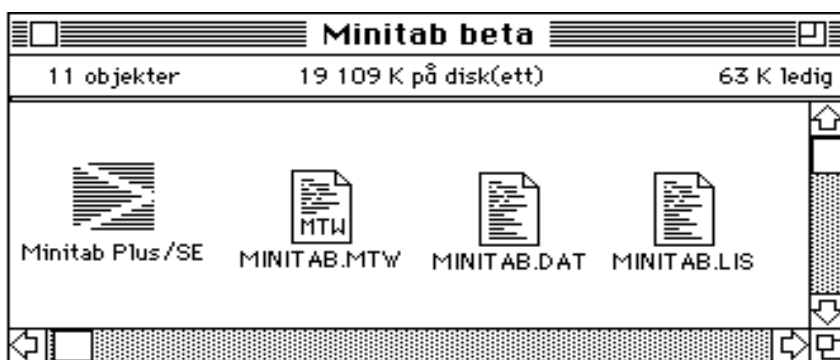
Du setter da inn den Minitab-disketten Minitab spør etter og trykker Enter (returtasten). Hvis du skriver et punktum, avslutter du Minitab uten at du får sjansen til å lagre det du har holdt på med.

Grunnen til dette diskettbytingssirkuset er at Minitab er et så stort program at alt ikke får plass i internhukommelsen samtidig. Programmet er derfor delt opp (med en teknikk som i dataspråk kalles "overlays") slik at kommandoene ligger på forskjellige disketter. Hvis du gir en kommando som ikke befinner seg på den disketten du har inne i diskettstasjonen, vil Minitab spørre deg etter den. På en harddisk er jo alle diskettene kopiert til et område på harddisken — dermed vil Minitab finne

kommandoene sine uten at det er nødvendig for deg å gjøre noen (du vil ikke se det engang).

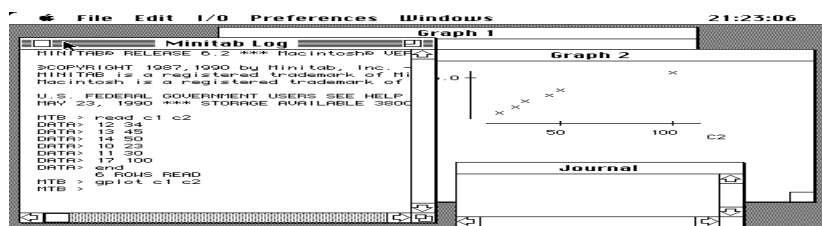
Minitab på Apple Macintosh

zzzz NY VERSJON Minitab for Macintosh finnes når denne manualen skrives (april 1990) kun som en Beta-versjon. Dette vil si at det er en test-versjon som kan inneholde feil, fordi den ikke er ferdig utviklet ennå. En ferdig versjon av Minitab for Macintosh er ventet i slutten av 1990 — men vær oppmerksom på at forsinkelser er svært vanlig i software-bransjen. Denne versjonen vil, som Beta-versjonen, tilsvare en 6.1 versjon for PC.



Minitab kan kjøres på alle Macintosh-modeller fra MacPlus og SE oppover til den foreløpige toppmodellene Mac IIfx. En harddisk er nødvendig. Det er to versjoner av programmet: en versjon for Mac Plus og SE, og en versjon for Mac II og andre "tyngre" Macintosh-er.

Du starter Minitab ved å "dobbelklikke" på Minitab-symbolet. Programmet ser ut omtrent som Minitab for IBM PC — dvs. at du gir kommandoer (og ikke bruker "pek-og-klikk" som med andre Macintosh-programmer). Endel kommandoer finnes på en rullemeny — det er alle kommandoer som



har med filer å gjøre (SAVE/RETRieve, WRITE/READ, OUTFile/NOOUTfile etc.). Det er imidlertid mulig å gi også disse kommandoene på samme måte som spesifisert i denne manualen — og av og til er det nødvendig, f.eks. når du skal lese inn i bestemte kolonner eller rader (SET- eller READ-kommandoen), eller når du skal benytte kommandoene uten filnavn (f.eks. READ C1 C2).

7.29 Spesielle kommandoer (Macintosh)

Innenfor Minitab for Macintosh finnes det endel "småting" som er spesielle for denne datamaskinen. Blant disse finner du

- *scrolling* ("screen rolling") av utskriften. Dette vil si at du kan bla deg tilbake i utskriften fra det du tidligere har gjort i Minitab ved å bruke pil og mus på samme måte som du blar deg opp og ned i tekstbehandlingsprogrammer (dvs. ved å flytte firkanten på den høyre vindusrammen opp og ned).
- *journalen* (dvs. resultatet av JOURNAL-kommandoen) kan du ha fremme på skjermen i et eget vindu på samme tid som du gir kommandoene.
- du kan bruke *klipp-og-lim funksjonen* (cut and paste) til å "hente" tidligere utførte kommandoer slik at du slipper å skrive dem om igjen. Dette kan du gjøre ved å "sverte" den tidligere kommandoteksten, velge "Cut" eller "Copy" fra Edit-menyen, gå ned til slutten av utskriften og velge "Paste" fra Edit-menyen. Kommandoen vil da bli lagt inn på kommandolinjen — og utføres, evt. med dine endringer, når du trykker Enter-tasten.

7.30 Filnavn og -typer (Macintosh)

Minitab sørger selv for å lage riktig filtype for de spesielle filene som brukes. Disse filtypene er

.MTW for Minitab worksheets (lagres med SAVE-kommandoen, hentes med RETRIEVE-kommandoen). Hvis du ikke spesifiserer noe filnavn i det hele tatt, bruker Minitab filnavnet MINITAB.MTW.

.DAT for redigerbare datafiler (datafiler som hentes og lagres med henholdsvis READ- og WRITE-kommandoen).

.LIS for OUTFile-filer. OUTFile-kommandoen krever et filnavn — så noe "standard" filnavn finnes ikke.

.MTB for kommandofiler. Dette gjelder kommandofiler som lagres ved hjelp av STORE-kommandoen, og kjøres ved hjelp av EXECUTE-kommandoen. Hvis du ikke bruker noe filnavn (typisk ved små kommandoløkker) bruker Minitab filbetegnelsen MINITAB.MTB.

.MTJ for JOURNAL-filer (filer laget med JOURNAL-kommandoen). Hvis du ikke spesifiserer noe filnavn, bruker Minitab filnavnet MINITAB.MTJ. **NB!** EXECUTE-kommandoen ser etter filer med type .MTB hvis du ikke spesifiserer noe annet — du må derfor gi filen nytt navn (filtype) hvis du ønsker å slippe å spesifisere filtypen.

På Minitab for Macintosh finnes ikke muligheten for å ha en oppstartkommandofil (slik som STARTUP.MTB på IBM PC).

7.31 Kapasitetsbegrensninger (Macintosh)

Størrelse på worksheet:	38.000 celler. Imidlertid: antall celler du kan benytte er endel mindre i praksis, fordi Minitab trenger ledige celler til mellomlagring ved beregninger, sortering og annet.
Max. antall kolonner:	1000
Max. antall konstanter:	1000

Max. antall matriser:	15
OW (Output Width):	79 tegn i bredden (maksimalt 132)
OH (Output Height):	ikke aktuell (Mac'en har vinduer som lar deg "rulle" teksten tilbake på samme måte som tekstbehandling)
IW (Input Width):	160 tegn (maks. bredde på datafiler).
BRIEF:	2 (utskrift av resultat av REGress og endel andre kommandoer)
HEIGHT:	17 linjer (høyden på PLOT)
WIDTH:	57 tegn (bredden på PLOT)
Maks. antall åpne filer:	6 (dette vil si at du kan ha 6 kommandofil-lag (5 hvis du har en OUTFILE åpen)).

7.32 Grafikk på Macintosh

Du har på samme måte som for IBM PC muligheten til å få "forbedret" grafikk hvis du gir de vanlige grafikkommandoene med en "G" foran (f.eks. blir PLOT til GPLOT). Minitab lager da et nytt "vindu" med grafen i. Du kan ha inntil 15 slike vinduer oppe på skjermen samtidig — hvis du forsøker å lage fler går det greit, men de "eldste" blir slettet. Hvis du forandrer størrelsen på et grafvindu, vil alle grafer som blir laget etter det tilpasset den nye vindusstørrelsen.

Minitab på IBM stormaskin (VM/CMS)

Minitab for stormaskin er ikke lenger et "grunnleggende" verktøy — stormaskinens kapasitet hva gjelder antall data og regnehastighet gjør at du kan benytte Minitab til mer enn opplæring — du kan virkelig svinge deg med makroprogrammering og kjempestore worksheets. Den aktuelle versjon av Minitab for VM/CMS er 7.1 idet denne manualen går i trykken.

Minitab startes ved å skrive

MINI TAB

på kommandolinjen. En skjerm med opplysninger, heading osv. for Minitab kommer frem. Ved å bruke CLEAR-tasten (på eldre terminaler: ALT + CLEAR) fjernes dette skjermbildet, og Minitab's klartegn (*MTB >*) kommer til syne.

7.33 Lagringsplass og andre standard

Antall dataelementer som kan lagres blir oppgitt idet du starter opp Minitab. På skjermen vil du få f.eks.

STORAGE AVAILABLE 229294

Dette betyr at du kan ha inntil 229294 tall lagret, men det er det samme om du har disse lagret som elementer i en kolonne, konstanter eller elementer i en matrise.

Lagringsplassen avhenger av størrelsen på ens virtuelle maskin (eller, for å si det litt upresist, den internhukommelsen du er tildelt). Denne størrelsen kan økes innenfor visse grenser av brukeren selv. Imidlertid er det vel de færreste som har behov for å benytte over 200.000 dataelementer — i så fall bør du nok benytte andre statistikkverktøy enn Minitab.

Størrelse på worksheet: Avhenger av størrelsen på ens virtuelle maskin (se ovenfor)

Max. antall kolonner: 1000

Max. antall konstanter: 1000

Max. antall matriser: 100

OW (Output Width): 79 tegn i bredden

OH (Output Height): 21 linjer (satt til 150 på BI)

IW (Input Width): 80 tegn (maks. bredde på datafiler). **NB!** Data- og kommandofiler må ha **fast** (fixed) format.

BRIEF: 1 (utskrift av resultat av REGress og endel andre kommandoer)

HEIGHT: 16 linjer (høyden på PLOT)

WIDTH: 50 tegn (bredden på PLOT)

Maks. antall åpne filer: 5 (dette vil si at du kan ha 5 kommandofil-lag (4 hvis du har en OUTFILE åpen).

Spesielle tegn: Bruk "Æ" (stor Æ) i stedet for "#" (amerikansk nummertegn) som kommentartegn.

7.34 Filnavn og -typer (IBM stormaskin)

Minitab sørger selv for å lage riktig filtype for de spesielle filene som brukes. Disse filtypene er

MINITABW	for Minitab worksheets (lagres med SAVE-kommandoen, hentes med RETRIEve-kommandoen). Hvis du ikke spesifiserer noe filnavn i det hele tatt, bruker Minitab filnavnet MINITAB WRKSHEET A.
DATA	for redigerbare datafiler (datafiler som hentes og lagres med henholdsvis READ- og WRITE-kommandoen).
LISTING	for OUTFILE-filer. Denne kommandoen krever et filnavn. Hvis du i stedet benytter PAPER-kommandoen er filnavnet MINITAB LISTING A (se avsnitt om PAPER-kommandoen).
MINITABC	for kommandofiler. Dette gjelder kommandofiler som lagres ved hjelp av STORE-kommandoen, og kjøres ved hjelp av EXECute-kommandoen. Hvis du ikke bruker noe filnavn (typisk ved små kommandoløkker) bruker Minitab filbetegnelsen MINITAB COMMANDS A.
MINITABJ	for JOURNAL-filer (filer laget med JOURnal-kommandoen). Hvis du ikke spesifiserer noe filnavn, bruker Minitab filnavnet JOURNAL MINITAB A.

Det anbefales å bruke disse filtypene for de ulike Minitab-filene. Det er imidlertid mulig å bruke andre filtyper, men da må både filnavn og filtype spesifiseres i den aktuelle Minitab-kommando.

CMS søker først etter filer som ligger på A-disken, deretter på B-disken osv. Derfor er det kun nødvendig å spesifisere file mode ("disk") hvis du skal ha tak i filer som også finnes på en "lavere" disk (f.eks. hvis du skal ha tak i en fil som ligger på X-disken, og har samme filnavn som en på A-disken).

På Minitab for stormaskin kan du ha en kommandofil som blir kjørt automatisk idet du starter Minitab (slik som STARTUP.MTB på IBM PC). Den må hete STARTUP MINITABC.

7.35 PAPER-kommandoen (IBM stormaskin)

PAPER-kommandoen er ikke implementert skikkelig under VM/CMS — det anbefales å benytte OUTFile-kommandoen.

Hvis du skriver ut en OUTFile (som får filtype LISTING hvis du ikke spesifiserer noe annet), vil det bli foretatt sideskift svært hyppig (fordi filen har såkalt FORTRAN-utskriftformat). Hvis du ikke ønsker dette kan du gå inn på filen i Xedit og fjerne den første kolonnen (se Xedit-manual). Du kan også gi filen en annen filtype enn LISTING — men da vil det komme et 1-tall i venstre marg hver gang Minitab har satt inn et sideskift.

7.36 CMS subset

Ved å bruke kommandoen

CMS

kommer du ut av Minitab og direkte over i CMS subset. Ved å skrive

RETURN

kommer du tilbake der du var i Minitab igjen.

Når du er i CMS SUBSET, kan du bruke de fleste CMS-kommandoer, inklusive FILELIST og systemeditoren XEDIT. Du kan bl.a. bruke dette for å kjøre batch-filer, observere resultatet, gå ut av Minitab, redigere batch-filen, for så å gå inn i Minitab igjen og kjøre batch-filen på nytt. Hvis du i tillegg programmerer inn CMS- og RETURN-kommandoen på noen PF-taster, vil du ha et både kjapt og elegant verktøy for store og/eller kompliserte statistiske beregninger. (Dette kan du gjøre ved å gi kommandoene SET PF9 IMMED CMS og SET PF10 IMMED RETURN før du starter Minitab).

7.37 Utskrift på skjermen (IBM stormaskin)

Når en skjerm er "full", står det MORE. . . nede i høyre hjørne av skjermen. For å se neste skjerm må du da trykke CLEAR (på eldre terminaler ALT + CLEAR). Hvis du har skrevet ut et eller annet som går over flere skjermer, stopper Minitab etter 150 linjer og spør

Conti nue?

Du svarer da Yes eller No og trykker Enter-tasten. (Eller, for å si det litt mer "profesjonelt": OH er satt til 150 linjer).

Minitab på DEC VAX (VMS)

Minitab på DEC (Digital Electronic Corporation) VAX²⁶ zzzz minner på mange måter om Minitab på IBM PC. Forskjellen ligger i at det ikke finnes noen fullskjerms dataeditor, at du har større frihet hva gjelder valg av filnavn, samt at kapasiteten er adskillig større — som på en IBM stormaskin er det kun maskinen som setter grenser, ikke Minitab. Operativsystemet på VAX-serien er som regel VMS, og ligner svært på MS-DOS (det er f.eks. de samme kommandoene for å liste ut og slette filer). VMS kjennetegnes ved et dollartegn (\$) som klartegn (prompt) når maskinen er klar til å utføre en kommando.

Minitab startes ved å skrive

MI NI TAB

på kommandolinjen.

7.38 Lagringsplass og andre standarder

Antall dataelementer som kan lagres blir oppgitt idet du starter opp Minitab. På skjermen vil du få f.eks.

STORAGE AVAI LABLE 149276

Dette betyr at du kan ha inntil 149276 tall lagret, men det er det samme om du har disse lagret som elementer i en kolonne, konstanter eller elementer i en matrise. Lagringsplassen avhenger av størrelsen på den maskinen du har tilgang til, men skulle være tilstrekkelig for de bruksområder Minitab er tiiltenkt. Som tidligere nevnt — har du behov for å benytte over 200.000 dataelementer bør du nok benytte andre statistikkverktøy enn Minitab.

Andre kapasitetsdata:

Størrelse på worksheet: Avhenger av den enkelte maskinen (se ovenfor).

Max. antall kolonner: 1000

Max. antall konstanter: 1000

Max. antall matriser: 100

OW (Output Width): 79 tegn i bredden

OH (Output Height): 21 linjer

IW (Input Width): 80 tegn (maks. bredde på datafiler). **NB!** Data- og kommandofiler må ha **fast** (fixed) format.

BRIEF: 1 (utskrift av resultat av REGress og endel andre kommandoer)

HEIGHT: 16 linjer (høyden på PLOT)

WIDTH: 50 tegn (bredden på PLOT)

²⁶Når dette skrives (Januar 1993), er nettopp Minitab versjon 9 for VAX annonsert, men ikke distribuert ennå. Denne versjonen har meget gode grafikkmuligheter, særlig innenfor kvalitetkontroll. Den har også et nytt kommandospråk, som gjør at man kan foreta "skikkelig" programmering av statistiske undersøkelser, ikke bare kjeder av kommandoer.

Maks. antall åpne filer: 5 (dette vil si at du kan ha 5 kommandofil-lag (4 hvis du har en OUTFILE åpen).

7.39 Filnavn og -typer (DEC VAX)

Filnavn under VMS kan ha massevis av bokstaver i filnavnet; hvis du ikke benytter deg av muligheten til å dele opp et filnavn med punktum'er, kan du ha inntil 39 tegn. Hvis du deler filnavnet opp med punktum'er kan du ha 39 tegn mellom hvert punktum — problemet med ikke-deskriptive filnavn, som du ofte har med PC'er, er altså ikke aktuelt. . I tillegg putter VMS på et *versjonsnummer*; et nummer som forteller hvilken utgave av filen det er snakk om. Dette vil si at hvis du åpner en fil, endrer den, og lagrer den på nytt, vil den gamle utgaven av filen ikke bli slettet, men beholdes²⁷. I stedet får den nye utgaven et nytt versjonsnummer. Et eksempel: sett at du har et datasett som heter MINEDATA.MTW. Det fulle navnet²⁸ på denne filen er under VMS

```
MINEDATA.MTW;1
```

Hvis du leser dataene fra denne filen inn i Minitab med RETRIEve-kommandoen, endrer dem, og så lagrer med SAVE-kommandoen, vil du ha to datasett:

```
MINEDATA.MTW;1 (det "gamle" datasettet)
MINEDATA.MTW;2 (det "nye" datasettet)
```

Hvis du ikke spesifiserer noe versjonsnummer, bruker VMS den nyeste versjonen. Dette gjelder imidlertid ikke VMS' kommando for å slette filer (DEL) — da må du spesifisere versjonsnummeret, evt. bruke en stjerne (*) for å markere "alle versjoner". Dette vil bl.a. si at du ved å gjenta en OUTFILE-kommando med det samme filnavnet ikke "lagt til" det som kommer på skjermen til den nye OUTFile-filen; i stedet ender du opp med mange versjoner med det samme filnavnet, og må "klippe dem sammen" manuelt. Hvis du ønsker å rydde opp i filene dine (du vil få mange "gamle" versjoner etterhvert) kan du gi VMS-kommandoen PURGE, som sletter alle gamle versjoner, slik at du sitter igjen med de nyeste versjonene av alle dine filer.

Minitab sørger selv for å lage riktig filtype for de spesielle filene som brukes. Disse filtypene er

MTW	for Minitab worksheets (lagres med SAVE-kommandoen, hentes med RETRIEve-kommandoen). Hvis du ikke spesifiserer noe filnavn i det hele tatt, bruker Minitab filnavnet MINITAB.MTW.
DAT	for redigerbare datafiler (datafiler som hentes og lagres med henholdsvis READ- og WRITE-kommandoen).
LIS	for OUTFILE-filer.
MTB	for kommandofiler. Dette gjelder kommandofiler som lagres ved

²⁷Hvor mange versjoner som beholdes er avhengig av hva systemadministratoren har spesifisert for den enkelte maskin.

²⁸Dette forutsetter at filen finnes innenfor ditt eget *user directory*, dvs. på ditt eget lagringsområde i datamaskinen. En fullstendig identifikasjon av en fil under VMS kan f.eks. være

```
MINVAX::MINDISK:[MITT.DIRECTORY]MINEDATA.DAT;1
```

Filer i andres brukerområder kan leses og kopieres medmindre eieren har beskyttet dem.

hjelp av STORE-kommandoen, og kjøres ved hjelp av EXECute-kommandoen. Hvis du ikke bruker noe filnavn (typisk ved små kommandoløkker) bruker Minitab filbetegnelsen MINITAB.MTB.

MTJ

for JOURNAL-filer (filer laget med JOURnal-kommandoen). Hvis du ikke spesifiserer noe filnavn, bruker Minitab filnavnet MINITAB.MTJ.

Det anbefales å bruke disse filtypene for de ulike Minitab-filene. Det er imidlertid mulig å bruke andre filtyper, men da må både filnavn og filtype spesifiseres i den aktuelle Minitab-kommandoen (inklusive versjonsnummer).

På Minitab for VAX/VMS kan du ha en kommandofil som blir kjørt automatisk idet du starter Minitab. Den må, akkurat som på IBM-kompatible PC'er, hete STARTUP.MTB.

7.40 VMS-aksess

Ved å bruke kommandoen

```
MTB > SYSTem
```

kommer du ut av Minitab og direkte over til VMS' kommandonivå. VMS kjenner du igjen ved klartegnet (prompt'en) "\$". Ved å skrive

```
$ MI NI TAB
```

kommer du tilbake der du var i Minitab igjen.

Minitab og matriseregning

Matriseregning benyttes bl.a. til løsning av ligninger med flere ukjente (og forutsetter, naturligvis, at du har like mange ligninger som ukjente). Determinant- og matriseregning har utbredt anvendelse innen bl.a. mikroøkonomi, hvor man benytter såkalte Hessianske matriser (matriser av annen ordens deriverte) for å avgjøre hvorvidt ekstremal-punkter på en funksjon er maksima eller minima. I dette kapitlet vil jeg imidlertid holde meg til ligningsløsning — hensikten er å vise Minitab-kommandoer, ikke mikroøkonomi²⁹. Endel regneark har matrisefunksjoner, men disse er som regel mer kronglete og dårligere dokumentert enn Minitab's. (Grunnen til at Minitab har disse funksjonene, er dels at statistikere gjerne bruker endel matriseregning, dels at endel statistiske analyser (f.eks) regresjonsanalyse og principal component analysis³⁰) benytter matriseinvertering — dermed skal det ikke mye programmering til for å legge til matrisefunksjonene).

8.1 Hvordan løse ligninger med Minitab

Vi setter opp følgende ligningssett:

$$\begin{aligned}2x + 3y &= 13 \\x + y + z &= 0 \\5x - 6y + z &= -13\end{aligned}$$

Dette kan løses ved substitusjon — f.eks ved at du flytter det andre leddet i den første ligningen over på høyre side, får $x = 6,5 - 1,5y$, og så setter dette inn i de andre ligningene. Denne fremgangsmåten er grei nok, men blir vanskeligere når antallet ligninger blir større. Matriseregning er et annet alternativ: da setter du opp koeffisientene, de ukjente og på den høyre siden konstantleddene:

$$\begin{bmatrix} 2 & 3 & 0 \\ 0 & 0 & 0 \\ 5 & -1 & 0 \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 13 \\ 0 \\ -13 \end{bmatrix}$$

²⁹For en god oversikt over bruksområder, se Sydsæter (1978) og særlig Chiang (1984).

³⁰Se Afifi & Clark (1990), side zzzz for en god beskrivelse av hvordan dette foregår.

Hvis vi kaller den første matrisen A , den andre (som inneholder de ukjente) U , og matrisen (eller rettere sagt kolonnen) på høyre side av ligningssettet H , kan vi skrive dette

$$AU = H$$

$$U = H * A^{-1}$$

Ved å flytte koeffisientmatrisen over, og invertere den, får vi

Med andre ord: du kan få løsningen på en ligningssett direkte ved å uttrykke ligningssettet med matriser.

En liten hake ved dette er at det å invertere matriser kun med hjelp av blyant og papir tar utrolig lang tid. Jeg vil ikke beskrive fremgangsmåten her, men i stedet vise hvordan det kan gjøres fort og greit i Minitab:

```

MTB > READ 3 3 M1      (les inn koeffisientene i en 3x3 matrise)
DATA> 2 3 0           (den første ligningen)
DATA> 1 1 1
DATA> 5 -6 1
      3 ROWS READ      (Minitab har lest tallene inn i matrisen M1)
MTB > SET C1          (leser høyre side inn i kolonne C1)
DATA> 13 0 -13
DATA> END
MTB > INVERSE m1 m2   (Lager matrisen M2, som er M1 invertert)
MTB > MULT M2 C1 C2   (multipliserer den inverterte matrisen med høyresiden av
                       ligningene, putter resultatet i C2)

MTB > PRINT c2

      C2
      2    3    -5      (C2 inneholder nå løsningen)

```

Ligningssettet har løsningen $x = 2$, $y = 3$ og $z = -5$.

8.2 Matrisekommandoer

Matriser betegnes i Minitab med M1, M2 osv. (maksimalt antall varierer mellom de ulike datamaskintypene). For å opprette en matrise kan du bruke READ-kommandoen, f.eks. slik:

```

MTB > READ 3 3 M1
DATA> 16 36 0
DATA> 7 12 1
DATA> 0 0 12.
MTB >

```

Den generelle formen på READ-kommandoen er

READ *antall_rader* *antall_kolonner* *matrisenummer*

Du kan også opprette en matrise med COPY-kommandoen. Du må da dataene lagret i kolonner, og gir kommandoen

MTB > COPY C1-C3 M4

I dette eksemplet opprettes matrisen M4 med tre kolonner og like mange rader som det er rader i kolonnene. Hvis det er ulike antall rader i kolonnene, gir Minitab en feilmelding.

Her er en liste, med korte forklaringer, over andre matrisekommandoer:

DEFINE	lar deg lage en matrise av like tall.
PRINT	skriver ut en matrise på skjermen.
TRANSPOSE	snur matrisen slik at rader blir kolonner og omvendt. En annen måte å tenke seg dette er at matrisen blir "snudd" rundt hoveddiagonalen (rekken med tall fra øverste venstre til nederste høyre hjørne).
INVERT	Inverterer matrisen.
DIAGONAL	Denne kommandoen kan brukes på to måter: enten til å ta diagonalen (rekken med tall fra øverte venstre hjørne til nederste høyre hjørne i matrisen) og lage en kolonne, eller å ta en kolonne og lage en matrise med denne kolonnen som diagonal. I siste tilfelle blir de andre tallene in matrisen satt til 0.
EIGEN	gir eigenverdiene (også kalt latente røtter) og eigenvektorene til en matrise.
ADD (eller SUBTract)	legger til (eller trekker fra) et tall, en kolonne eller en matrise til (fra) en annen matrise. Det forutsettes at matrisene er like store.
MULTIPLY	ganger et tall, en kolonne eller en matrise med en matrise. NB! når du multipliserer matriser med hverandre, er rekkefølgen viktig. Matrisemultiplikasjoner er ikke kommutative, dvs. at $AB \neq BA$. Divisjon med en matrise er udefinert.

A. Kommentert litteraturliste

Minitab's hjelpefunksjon (HELP-kommandoen) er så god at du egentlig ikke trenger noen manual (utover denne, selvsagt. . . .). Her er imidlertid en oversikt over noe av den dokumentasjon som er å finne om Minitab:

Minitab Quick Reference (Minitab Inc.)

Dette er en "huskelapp", trykket på stiv kartong. Den følger med Minitab-utgaven for IBM PC og Macintosh, men kan brukes for alle utgaver av Minitab. Den lister opp alle kommandoene gruppert etter type, i stikkordsform.

Minitab Mini-Manual (Minitab Inc., July 1992)

Introduksjonsmanual til Minitab versjon 6, 7 eller 8 og fremover. Grei introduksjon, god oversikt over Minitab for IBM PC og Macintosh i de to siste kapitlene.

Minitab Reference Manual, Release 8 (Minitab Inc., november 1991)

Dette er det komplette og "offisielle" oppslagsverket om Minitab — som imidlertid kan oppleves som vanskelig hvis du ikke har kjennskap til datamaskiner og statistikk. Her står absolutt alt du noensinne vil trenge å vite om Minitab. . . .

Minitab Handbook for Business and Economics (Robert B. Miller, PWS-Kent Publishing Company, Boston, USA., 1988)

En lærebok i Minitab spesielt beregnet på økonomisk-administrative problemstillinger. Mange gode eksempler.

Minitab Handbook, Second Edititon (Ryan/Joiner/Ryan, PWS Publishers, Duxbury Press, Boston, USA., 1985)

Meget god nybegynnerbok med masse eksempler. Siste utgave er oppdatert t.o.m. versjon 6.1.1 av Minitab. Muligens noe unødvendig siden Minitab Mini-Manual kom ut.

Statistikkbøker som det er referert til i teksten (eller som er brukt som grunnlag for denne manualen):

Afifi & Clark zzzz

Denne boken er en statistikkbok for folk som vil drive forskning innenfor sosialfag. Den beskriver meget godt analysemetoder for spørreskjema, som f.eks. LOGIT regresjon, transformasjoner, PCA og andre multivariate metoder.

Ben-Horim, M. of H. Levy: *Statistics, Decisions and Applications in Business and Economics*, 2. utgave. Random House, New York, NY 1984.

Grunnleggende statistikkbok med fin disposisjon. Denne boken benyttes på siviløkonomstudiet på BI.

Berenson, Mark L. og David M. Levine: *Basic Business Statistics, Concepts and Applications*, Prentice-Hall Inc., Englewood Cliffs, NJ., 1979.

Grunnleggende statistikkbok, lettlest. Denne ble tidligere benyttet på siviløkonom- og revisorstudiet på BI.

Hair Jr., Joseph F., Ralph E. Anderson og Ronald L. Tatham (1987): *Multivariate Data Analysis with Readings*, 2nd edition, Collier Macmillan Publishers, London.

Lettlest (men ordrik) bok om multivariate metoder. Her får du en god beskrivelse av multipel regresjonsanalyse og diskriminantanalyse. Tar sine eksempler hovedsaklig fra markedsanalyser.

Tufte, Edward: *The Visual Display of Quantitative Information*, 1983, og *Envisioning Information*, 1990. Begge: Cheshire, Connecticut: Graphics Press.

Den første av disse to bøkene er en klassiker; den handler om hvordan du skal og ikke skal bruke grafikk for å fremstille data. Tufte viser gode og dårlige eksempler på bruk av ulike typer diagrammer (søylediagrammer, kakediagrammer o.l.), samt hvordan du kan få data til å se ut som noe det ikke er ved å benytte ulike typer grafikk. Glimrende skrevet og rikt illustrert, en underholdende bok som bør leses av alle som skriver eller leser rapporter eller tidsskriftartikler som inneholder grafikk. Bl.a. tar Tufte en gang for alle livet av regnearkenes fancy "business graphics"; han viser hvordan du ved å legge på mer og mer farger og dimensjoner kan forvirre leseren til å tro at dataene forteller en annen og mer interessant historie enn de egentlig gjør. . . . Den andre boken tar for seg mer avansert datagrafikk, og forsøker å finne nye måter å fremstille informasjon på, slik at den blir interessant uten at du fordreier sannheten eller forvirrer leseren.

Andre referanser:

Berk, Richard A.: "Applications of the General Linear Model to Survey Data", in Rossi, P.H., Wright, J.D. and Anderson, A.B. (eds.): *Handbook of Survey Research*, San Diego, California: Academic Press, 1983

Boyer, Carl B.: *A History of Mathematics*, 2. utgave, New York: Wiley 1991.

Campbell, Jeremy: *Grammatical Man*, New York: Simon and Schuster, 1982

Chiang, Alpha C.: *Fundamental Methods of Mathematical Economics*, 3. utgave, New York: McGraw-Hill 1984.

Campbell, D.T. and Stanley, J.C.: *Experimental and Quasi-experimental Designs for Research*, Houghton Mifflin Company, 1963.

Cook, T.D. and Campbell, D.T.: *Quasi-Experimentation*, Chicago, Rand-McNally, 1977.

"OMNITAB on the 90", *Datamation*, Vol. 9, No. 3, Mars 1963.

Hanushek, Eric A. and Jackson, John E.: *Statistical Methods for Social Scientists*, Orlando, Florida: Academic Press, 1977.

- Hilsenrath, J. et al., "OMNITAB — A Computer Program for Statistical and Numerical Analysis".
National Bureau of Standards Handbook 101, Washington, D.C.: National Bureau of Standards, 1966.
- Huff, Darrell: *How to Lie with Statistics*, New York: W.W: Norton & Co., 1954.
- Lapin, Lawrence L.: *Statistics for Modern Business Decisions*, Fourth Edition, Orlando, Florida: Harcourt Brace, 1987.
- Raskin, Robin: "Statistical Software for the PC; Testing for Significance", *PC Magazine*, No. 5 1989, New York: Ziff-Davis Publishing Company
- Ryan Jr., T. A. og B. L. Joiner: "Normal Probability Plots and Tests for Normality", Technical Report, Statistics Department, The Pennsylvania State University, 1976.
- Sammet, Jean: *Programming Languages: History and Fundamentals*, Englewood Cliffs, N.J.: Prentice-Hall Inc., 1969.
- Sydsæter, Knut: *Matematisk Analyse*, bind 2, Universitetsforlaget, 1978.
- Tversky, Amos & Daniel Kahnemann: "Judgment under Uncertainty: Heuristics and Biases", *Science*, September 1974, pp. 1124-1131
- Tversky, Amos & Daniel Kahnemann: "The Framing of Decisions and the Psychology of Choice", *Science*, January 1981, pp. 453-458
- Wenstøp, Fred: *Statistikk og dataanalyse*, Tano, Oslo 1987

B. Data brukt i eksempler

Her er en utlisting av "studentdataene" som er benyttet i de fleste eksemplene i manualen.

Noen tilleggsopplysninger:

- dataene stammer fra et spørreskjema som ble delt ut blant siviløkonomstudentene i 1. årskull på BI høsten 1983. De er benyttet endel i statistikkundervisningen på BI.
- "Sigarett" er antall sigaretter en person røyker pr. uke.
- "Royking" er konstruert ved hjelp av CODE-kommandoen, og er 1 hvis en person har et resultat ≥ 1 i C6 "Sigarett".
- i kolonnen for kjønn ("Kjonn") er 1 = mann, 2 = kvinne.

C1 Rad	C2 Kj - onn	C3 Hoy- de	C4 Vekt	C5 Sko- nr.	C6 Al - der	C7 Si ga- rett	C8 Roy- king	C9 C1	C10 C2	C11 C3	C12 C4	C13 C5	C14 C6	C15 C7	
1	2	176	70.0	38.5	20	100	1	51	1	183	92.0	43.5	21	0	0
2	1	172	63.0	41.0	21	0	0	52	1	190	82.0	45.0	23	0	0
3	1	170	72.0	41.0	20	0	0	53	1	182	74.0	43.0	21	0	0
4	1	174	69.0	40.0	20	0	0	54	1	182	70.0	41.5	22	0	0
5	2	166	52.0	38.0	20	0	0	55	1	181	72.0	43.0	22	0	0
6	1	194	87.0	45.0	21	0	0	56	1	189	91.0	46.0	21	0	0
7	1	179	67.0	42.0	20	0	0	57	1	186	80.0	43.0	21	5	1
8	1	173	69.0	42.0	22	80	1	58	1	184	83.0	44.0	23	0	0
9	1	180	70.0	42.0	21	5	1	59	1	190	90.0	44.0	20	0	0
10	1	180	74.0	43.0	21	0	0	60	1	171	71.5	41.0	21	0	0
11	2	183	57.0	40.0	20	0	0	61	1	182	73.0	43.0	21	0	0
12	1	190	77.0	44.0	21	0	0	62	1	196	73.0	45.0	21	50	1
13	1	177	68.0	43.0	20	0	0	63	1	175	71.0	41.5	20	7	1
14	1	180	72.0	42.0	22	0	0	64	1	183	62.0	43.0	20	0	0
15	2	175	65.0	38.0	20	10	1	65	1	192	88.0	44.0	22	0	0
16	2	178	63.0	38.0	21	70	1	66	1	193	80.0	43.0	21	0	0
17	1	184	78.5	42.0	24	0	0	67	1	186	85.0	43.0	21	0	0
18	2	166	57.0	38.0	22	3	1	68	1	179	64.0	41.0	22	0	0
19	2	170	52.0	38.0	22	10	1	69	1	183	73.0	43.0	21	0	0
20	1	170	65.0	41.0	23	0	0	70	1	177	67.0	42.0	20	0	0
21	2	169	52.0	38.0	20	0	0	71	1	179	65.0	43.0	20	0	0
22	1	178	58.0	41.0	20	0	0	72	2	168	58.0	39.0	20	1	1
23	2	169	52.0	37.5	21	0	0	73	1	187	80.0	42.0	21	0	0
24	1	187	71.0	43.0	22	100	1	74	1	179	65.0	39.0	21	0	0
25	1	180	70.0	42.0	24	0	0	75	1	172	63.0	42.0	22	0	0
26	1	192	82.0	46.0	21	0	0	76	1	188	73.0	42.0	22	0	0
27	1	184	85.0	43.5	22	50	1	77	2	170	53.0	37.0	23	0	0
28	1	190	74.0	44.0	21	50	1	78	1	180	72.0	42.0	21	0	0
29	1	179	67.0	43.0	22	4	1	79	2	163	55.0	37.0	20	0	0
30	1	192	82.0	44.0	20	10	1	80	1	188	90.0	43.0	21	3	1
31	1	184	79.0	44.0	24	0	0	81	1	178	70.0	42.0	21	0	0
32	1	190	80.0	45.0	22	0	0	82	1	190	83.0	44.5	21	0	0
33	1	178	67.0	41.0	21	5	1	83	1	183	68.0	41.5	22	0	0
34	1	185	73.0	43.5	20	0	0	84	1	171	59.0	40.0	19	0	0
35	1	186	67.0	43.0	23	0	0	85	1	189	73.0	42.0	22	0	0
36	1	174	71.0	43.0	22	*	*	86	1	186	93.0	44.0	23	70	1
37	1	175	70.0	41.0	22	0	0	87	1	189	83.0	43.5	22	0	0
38	1	187	82.0	43.0	20	0	0	88	1	180	73.0	42.0	24	0	0
39	1	183	82.0	44.0	20	20	1	89	1	183	75.0	41.0	22	0	0
40	1	185	80.0	45.0	20	0	0	90	1	186	75.0	43.0	20	0	0
41	1	183	80.0	46.0	21	0	0	91	1	170	59.0	41.0	20	0	0
42	1	180	60.0	40.0	22	0	0	92	1	177	69.0	43.0	21	0	0
43	1	180	65.0	43.0	20	0	0	93	1	190	75.0	43.0	21	0	0
44	2	172	60.0	38.5	20	0	0	94	1	182	74.0	41.5	20	0	0
45	1	185	69.0	42.0	20	0	0	95	1	186	80.0	44.0	19	0	0
46	2	172	58.0	38.0	20	0	0	96	1	179	71.0	41.0	23	0	0
47	1	190	77.0	45.0	20	0	0	97	1	195	83.0	43.0	21	0	0
48	1	184	73.0	45.0	20	0	0	98	1	171	70.0	42.0	23	175	1
49	1	170	56.0	38.0	22	0	0	99	1	177	74.0	44.0	22	0	0
50	1	186	70.0	44.0	19	0	0	100	1	182	73.0	42.5	22	0	0

Rad	C1 Kj- onn	C2 Hoy- de	C3 Vekt	C4 Sko- nr.	C5 Al- der	C6 Si ga- rett	C7 Roy- king
101	2	171	56.0	39.0	23	140	1
102	2	168	55.0	38.0	22	15	1
103	1	174	62.0	41.0	24	0	0
104	1	172	69.0	40.0	25	0	0
105	1	186	70.0	42.5	21	2	1
106	1	182	70.0	42.5	21	0	0
107	1	168	60.0	41.0	21	0	0
108	1	181	73.0	43.0	22	70	1
109	1	190	78.0	44.0	21	0	0
110	1	189	77.0	43.5	20	7	1
111	1	180	78.0	43.0	36	0	0
112	2	167	63.0	39.0	24	110	1
113	1	193	87.0	43.5	23	*	*
114	1	183	69.0	43.0	25	140	1
115	1	179	77.0	41.5	22	30	1
116	1	177	71.0	39.0	20	0	0
117	1	172	70.0	42.0	19	0	0
118	1	188	83.0	43.5	20	30	1
119	1	176	70.0	42.0	22	60	1
120	1	183	67.0	42.0	21	0	0
121	1	189	82.0	44.0	21	0	0
122	2	178	62.0	40.0	20	30	1
123	1	185	80.0	44.0	22	40	1
124	1	179	69.0	43.0	21	0	0
125	2	169	65.0	38.0	20	5	1
126	2	174	62.0	39.0	21	0	0
127	2	170	55.0	37.0	26	0	0
128	1	183	82.0	43.0	24	110	1
129	1	182	78.0	42.5	24	0	0
130	1	180	74.0	42.5	21	0	0
131	1	170	58.0	39.0	22	10	1
132	1	173	74.0	43.0	20	0	0
133	1	175	71.0	41.0	21	30	1
134	1	181	70.0	43.0	22	0	0
135	1	176	69.0	42.0	21	0	0
136	1	175	72.0	42.0	22	100	1

137	1	184	88.0	44.0	22	140	1
138	1	182	67.0	42.0	21	0	0
139	1	184	76.0	45.0	22	4	1
140	1	190	76.0	43.0	21	0	0
141	1	189	80.0	43.0	22	0	0
142	1	185	77.0	43.0	23	100	1
143	1	185	72.0	43.0	21	0	0
144	1	189	72.0	42.5	21	0	0
145	1	186	77.0	42.0	21	0	0
146	1	178	65.0	40.0	21	0	0
147	1	188	78.0	44.0	21	0	0
148	1	176	63.0	41.0	22	*	*
149	1	174	70.0	43.0	25	0	0
150	2	179	67.0	40.0	19	0	0
151	2	171	55.5	38.0	21	0	0
152	2	168	64.0	38.0	20	5	1
153	1	195	89.0	46.0	22	0	0
154	2	163	53.0	39.0	21	8	1
155	1	190	83.0	45.0	22	*	*
156	1	184	76.0	43.5	25	140	1
157	1	187	72.0	42.0	21	0	0
158	1	187	76.0	44.0	20	2	1
159	1	170	59.0	41.0	22	0	0
160	1	189	87.0	45.0	22	0	0
161	2	166	49.0	37.0	20	0	0
162	1	183	72.0	41.0	22	0	0
163	1	184	80.0	43.0	21	0	0
164	1	195	92.0	45.0	21	0	0
165	1	178	75.0	40.0	23	0	0
166	2	164	51.0	38.0	23	0	0
167	1	183	81.0	43.0	20	0	0
168	2	164	55.0	37.5	21	0	0
169	2	164	47.0	36.5	21	0	0
170	2	173	56.0	37.0	25	0	0
171	1	187	83.0	43.5	21	0	0
172	2	170	66.0	37.0	21	0	0
173	1	186	78.0	45.0	22	0	0
174	1	182	68.0	42.0	24	210	1
175	2	167	55.0	37.0	20	0	0
176	1	173	57.0	40.0	21	0	0
177	1	190	76.0	45.0	21	0	0

Følgende data, som viser levetiden for bildekk i 1000 miles produsert på tre forskjellig skift, brukes i beskrivelsen av "oneway" vari-ansanalyse:

C1 Day	C2 Eveni ng	C3 Ni ght
25. 40	23. 40	20. 00
26. 31	21. 80	22. 20
24. 10	23. 50	19. 75
23. 74	22. 75	20. 60
25. 10	21. 60	20. 40

Dataene er hentet fra Berenson & Levine (1979), side 388.

D. Stikkordregister

- * (manglende observasjon) 16
- # (kommentartegn) 15
- & (fortsettelse av linje) 15
- ABORT (avbryt kommando) 15
- ABSOLute (absoluttverdi) 36
- Absoluttverdi 36
- ACOS (Arccosinus) 38
- ALTErnative (subkommando) 63, 65-67, 99
- Anderson, Ralph E. 138
- ANTIlog 36
- AOVOneWay 69, 70
- Arccosinus (ACOS) 38
- Arcsinus (ASIN) 38, 38
- Arctangens (ATAN) 38
- ASCII 106, 115
- ASIN (Arcsinus) 38
- Asterisk 16
- ATAN (Arctangens) 38
- Avbryte kommando 15
- BASE (startpunkt RANDOM) 56, 58
- BATCh 105, 108, 110, 130
- Ben-Horim, M. 137
- Berenson, Mark L. 69, 86, 138, 141
- Berk, Richard A. 85, 86, 88, 138
- BOXPlot 9, 72, 75-77
- Boyer, Carl B. 51, 138
- BRIEf 82, 103, 103, 107, 120, 127, 128, 132
- Campbell, Donald T. 84
- Campbell, Jeremy 138
- CDF (Cumulative Distribution Function) 51, 53-56
- Chiang, Alpha C. 134, 138
- CMS (operativsystem IBM stormaskin) 7, 128, 129
- CMS SUBSET 130
- CODE 17, 32, 36, 140
- COLPercents (subkommando til TABLE) 47, 109
- COMMands (subkommando til HELP) 17, 18, 72, 129
- Constants 11, 13, 17, 32
- Continue? 18, 103, 108, 108, 130
- CONVert 26
- Cook, Thomas D. 84, 139
- COPY 14-16, 29-30, 39, 59, 110, 126, 136
Subkommandoer 14, 15, 29-30, 93
- CORRelation 85, 87
- COS (Cosinus) 38
- Cosinus 38
- COUNt 32, 38, 41, 42-45, 49, 57, 68, 74, 75, 96
- COUNTs (subkommando til TABLE) 46, 69
- CPU 11
- Dataeditor 28, 119, 131
- Datamation 139
- DESCRibe 41, 42, 109, 111, 111
- DISCReTE (sanns.ford.) 51, 52, 60, 110
- DOS (operativsystem PC) 6, 24, 107, 115, 116, 118, 121, 131
- DOTPlot 58, 72, 76, 77, 100, 111
- ECHO 109-110
- END 14, 15, 20-22, 24, 60, 68, 106-108, 110, 135
- ERASe 31
- EXECute (kjør kommandofil) 106-111, 120, 126, 126, 129, 133
- EXIT (DOS-kommando) 121
- EXPOnentiate 36
- FORMat 22-25, 70, 105, 113-115, 128, 131
- Fortegnstest 99
- FORTRAN-utskriftformat 9, 25, 113, 114, 130
- FREQUencies (subkommando til TABLE) 48, 49
- GBOXplot 121
- GHIStogram 121
- Gjennomsnitt 16, 38, 40, 41, 46, 62-65, 69, 85, 87, 104
- GOPTions 122
- GPLOt 121, 123, 127
- Grafikk 9, 72, 116, 118, 119, 121, 122, 127, 138
- Hair, Joseph F. 138
- Hanushek, Eric A. 93, 139
- Harddisk 7, 9, 12, 116, 118, 121, 123-125
- HELP 8, 17-19, 29, 53, 72, 76, 103, 137
COMMands 17, 18, 72, 129
- Heteroscedastisitet 84, 85, 87, 88
- Hilsenrath, J. 139

HISTogram 72, 74-77
 Histogrammer 72, 74-77
 Homoscedastisitet 85
 Huff, Darrell 139
 Hullkort 113
 IBM
 PC (og kompatible) 6, 23, 107, 118-120, 133
 Stormaskin 7, 10, 22, 72, 107, 109, 111, 128-131
 INFOrmation 13, 18, 44, 138
 INSErt 21, 22, 25, 114, 123
 INTEgers 10, 42, 52
 Interaktiv 13, 105, 110
 Internhukommelse 11, 12, 20, 25, 118, 118, 124, 128
 INVCDF 51, 56, 63-65, 68, 70
 IW (Input Width) 103, 120, 127, 128, 131
 Jackson, John E. 93, 139
 Joiner, B. L. 87, 137, 139
 JOURNal 105-107, 120, 126, 126, 129, 133
 Kahnemann, Daniel 139
 Kommandofil 105, 108-109, 115, 121, 127, 129, 129, 132, 133
 Kommandofiler
 å lage dem 106, 107
 å starte dem 106-111, 120, 126, 126, 129, 133
 Kommentarer 1, 15, 109, 142
 Konfidensintervall 16, 62, 64, 81, 82, 99
 Konstanter 7, 11, 13, 17, 22, 23, 28, 29-32, 82, 120, 127, 128, 131
 Krysstabulering 44
 Kumulative sannsynligheter 53-55
 Kvadratrot 36, 41
 Kvartil 41, 76
 LAG 37, 38, 109, 109, 121, 127, 129, 132
 Lapin, Lawrence L. 85, 139
 LAYOut (subkommando til Table) 49
 LET 32-35, 38, 60, 92, 106, 108, 110, 110
 Levels 109
 Levine, David L. 69, 86, 138, 141
 Levy, H. 137
 LINE (subkommando til GPLOT) 123
 LOGE (naturlige logaritme) 36
 LOTUS 1-2-3
 lagring av data 23, 24, 113
 Macintosh 1, 6, 7, 12, 22, 28, 34, 72, 106, 107, 111, 115, 125-127, 137
 Makro 9, 25, 110
 Maksimum 41
 Manglende verdi 11, 16, 36, 38, 46, 48
 MANN-Whitney 77, 100, 100
 Matriser 7, 11, 13, 22, 23, 28, 29, 31, 120, 127, 128, 131, 134-136
 MAXimum 41, 41, 111, 120, 127, 128, 131, 131
 MAXimums (subkommando til TABLE) 46
 MEAN (gjennomsnitt) 38, 41-42, 46, 47, 62-67, 70, 71, 111
 MEANs (subkommando til TABLE) 46
 MEDian 38, 41-43, 46, 75, 99-100, 111
 MEDians (subkommando til TABLE) 46
 Miller, Robert B. 137
 MINimum 38, 41, 43
 MINimums (subkommando til TABLE) 46
 Minitab
 Handbook 137-139
 Minitab Handbook 137
 Minitab Reference Manual 8, 92, 95, 137
 MISSing (subkommando til TABLE) 46
 Missing value 11, 16, 17, 41
 MSETUP 113, 122
 Multikolinearitet 84, 85, 97
 N* 41
 NAME 28, 29, 32, 44, 66, 87, 90, 92, 110
 Naturlige logaritme 36
 NMISs 38, 43
 NMISsing (subkommando til TABLE) 46
 NOALL (subkommando til TABLE) 48, 67
 NOConstant 82
 NOECho 108, 110
 NOJournal 107
 NOOoutfile 102, 116, 125
 NORMal 52, 55, 86, 87, 139
 Normalitet (forutsetning for regresjon) 84-86
 NOTE 15, 86, 109, 110, 123
 NOTErminial 102, 106, 108
 NSCOres 86, 87
 OH 31, 103, 120, 127, 128, 130, 131
 OMIT (subkommando til COPY) 29-30
 OMNITAB 7, 139
 ONEWay 9, 69, 70, 141
 Oppstartfil 108, 127, 130, 133

OUTFILE 18, 31, 33, 102-103, 106, 108, 115,
 116, 120, 121, 125-127, 129,
 130, 132, 132
 Outliers (ekstremverdier) 41, 75, 76, 81
 Output Height 103, 120, 127, 128, 131
 Output Width 103, 120, 127, 128, 131
 Overlays 124
 OVERview 17, 72
 OW 103, 120, 127, 128, 131
 PAPER 33, 102, 129, 130
 PARSums 36, 36
 PC 6, 23, 107, 118-120, 133
 PDF 51, 53, 54, 56
 PLOT 72, 73, 86, 87, 103, 120, 127, 129, 132
 POISSon 52, 53
 POOLed 65-67, 70, 71, 104
 PORTable (format) 22, 23, 105, 119
 PORTable (subkommando SAVE) 22, 23, 105,
 119
 PREDict 81, 82, 103
 PRINT 20, 26, 33, 40, 60, 114, 115, 135, 136
 Prompt 12, 13, 15, 20, 107, 121, 131
 PROPortion (subkommando til TABLE) 47
 Punktsannsynligheter 53
 Quick Reference 137
 RAM (internhukommelse) 11
 RANDOM 11, 18, 51, 56-60, 89, 110, 137
 Random Access Memory 11
 Rangering 37, 40
 RANK 37, 39, 40
 Raskin, Robin 139
 READ 20-25, 60, 68, 107, 110, 114, 115, 120,
 125, 125, 129, 133, 135, 136
 Regresjonsanalyse 1, 9, 16, 18, 72, 79, 80, 82,
 84-86, 88, 90, 91, 94, 134,
 138
 REGRes 80-82, 91, 92, 103, 104, 120, 127,
 128, 132
 REName (DOS-kommando) 120
 REPLace 59
 RESIDuals (subkommando til REGRes) 80,
 82, 86
 REStart 31
 RETRieve 22-24, 113, 119, 125, 126, 129,
 132, 133
 ROUND 36
 ROWPercents (subkommando til TABLE) 47,

109
 RUNS 98
 Ryan, T. A. Jr. 87, 137, 139
 SAME 58, 77, 100
 Sammet, Jean 139
 SAMPLe 56, 59
 SAVE 12, 13, 22-24, 105, 107, 113, 119-121,
 125, 126, 129, 132, 133
 LOTUs-format 23, 24, 113
 PORTable format 22, 23, 105, 119
 SEMEAN 41, 42, 111
 SET 8, 20, 21, 25, 107, 110, 114, 125, 130,
 135
 Sign test (STEST) 99
 SIGNs 36
 SIN (Sinus) 38
 Sinus 38
 SORT 6, 18, 37, 39, 39
 Sortering 6, 18, 37, 39, 39
 SPSSX (Statistikkpakke) 113
 SQRT 36, 41
 SSQ 38, 43
 STACk 31, 31
 Standardavvik 38, 41-43, 46, 62-67, 70, 71,
 80, 81, 91, 111
 Stanley, J. C. 139
 STARTUP (oppstartfil) 108, 127, 130, 133
 STATs (subkommando til TABLE) 46
 STDEv (standardavvik) 38, 41-43, 46, 62-67,
 70, 71, 80, 81, 91, 111
 STEM-and-leaf 77
 STEPwise (trinnvis multippel regresjon) 94
 STEST (Fortegnstest) 99
 STOP 12, 31, 116
 STORE (lagring kommandoer) 106, 107
 Subkommandoer 14, 18, 25, 29, 30, 39, 44-46,
 51, 65, 72, 76, 82, 95, 103
 avbrytelse av 15
 SUBScripts 32, 66
 SUM 35, 38, 43, 46, 60, 111
 SUMS (subkommando til TABLE) 46
 Sydsæter, Knut 134, 139
 SYSTEM 7, 13, 121, 133
 Tabell 45, 54, 56, 81, 95
 TABLE 19, 27, 44-49, 57, 67, 68, 106, 108,
 109
 TAN (Tangens) 38

Tangens 38
Tatham, Ronald L. 138
TINTerval 64
TOTPercents (subkommando til TABLE) 47
TRMEAN 41, 42, 111
TSHAre 108, 110
TTESt 64-67
Tuft, Edward 138
Tversky, Amos 139
TWOSample 65-67, 99
TWOT 67
UNIFORM (sanns.ford.) 52
UNSTack 32, 66
USE (subkommando til COPY) 14, 15, 29, 30,
93
Variansanalyse 69, 70
VM/CMS (operativsystem IBM stormaskin) 7,
128, 130
Wenstøp, Fred 58, 100, 101, 139
Worksheets 22, 23, 105, 113, 119, 126, 128,
129, 133
WRITE 12, 13, 23-25, 105, 114, 115, 120,
125, 126, 129, 133
ZINTerval 62
ZTESt 63, 65