Kort introduktion till `MINITAB`

Statistisk Kvalitetsstyrning

14 september 1995

Förord

Detta är en liten skrift som är tänkt att ge en snabb introduktion till `MINITAB`, version 9.1. Om du har hittat fel eller har synpunkter på hur man kan förbättra den får du gärna höra av dig till karin@math.chalmers.se eller nader@math.chalmers.se. Du kan också skicka brev till oss på adressen

Matematiska institutionen
CTH
412 96 GÖTEBORG

Karin Nelander och Nader Tajvidi

Innehåll

1 In- och utloggning

För att kunna köra `MINITAB` på datorn måste man först *logga in*, varvid man upplyser operativsystemet om vem man är. För att kunna logga in måste texten `login:` synas på skärmen[1]. Man skriver sedan sitt användarnamn och trycker på `<Return>` , se figur 1.

```
login:  stk95-18
password:
```

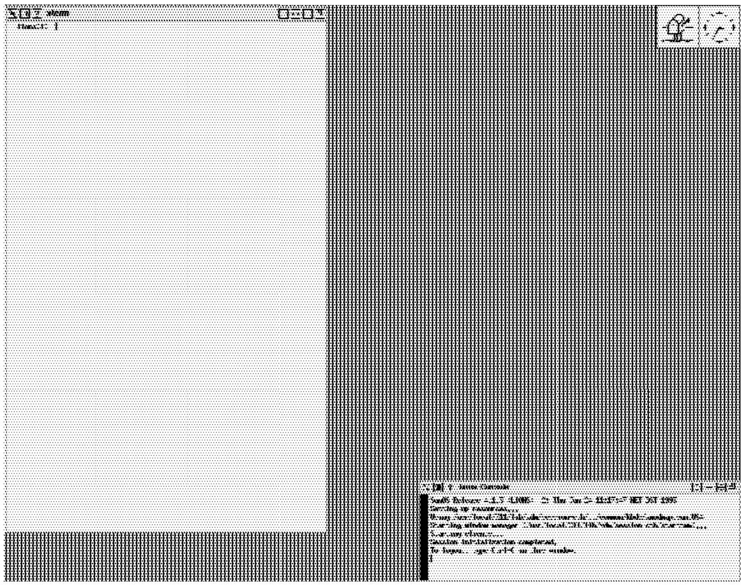
Figur 1: Att logga in.

Nästa fråga gäller ens personliga lösenord. När du skriver lösenordet visas ingenting på skärmen, detta för att ingen annan skall kunna läsa det.

Om allt går väl kommer det efter ett litet tag att se ut som i figur 2. Det stora fönstret till vänster är ett terminalfönster, det mindre fönstret längst ned är ett konsolfönster.

Dessutom finns några små fönster, t ex en klocka och en brevlåda. Det är det stora terminalfönstret som du skall jobba i.

Nästa steg är att starta `MINITAB`. Det gör du genom att placera musmarkören någonstans i det stora fönstret uppe till vänster. Här finns en så kallad *prompter* som talar om vilken dator du har loggat in på. Efter *promptern*, i detta fall `dali>`, skriver du mtb:



Figur 2: Bildskärmen efter start av X11.

`dali>` **mtb**

Efter att du gjort detta kommer terminalfönstret att se ut som i figur 3. I nästa avsnitt skall vi ge en kort genomgång av de kommandon i `MINITAB` som vi kommer att använda, men först kort om hur man loggar ut från systemet.

Gör så här:

- Skriv **STOP** i terminalfönstret. Då avslutas körningen av `MINITAB` .
  Varning: Nu försvinner allt man gjort som ej är sparat, se kommandona **SAVE** och **OUTFILE**.

- Placera musmarkören på bakgrunden och håll ner högra musknappen. Behåll musknappen nertryck och flytta musmarkören så att den sista raden (med texten `Avsluta`) i menyn blir markerad med svart. Släpp musknappen.

[1] Om denna text inte syns, flytta på musen.

Figur 3: Terminalfönstret efter start av `MINITAB` .

2 Introduktion till `MINITAB`

2.1 Inledning

`MINITAB` är ett statistiskt programpaket som kan användas till att organisera och analysera data, samt att rapportera resultat från statistiska analyser.

`MINITAB` kan delas in i två delar, minnesutrymmet och kommandona. Minnesutrymmet kan i sin tur delas in i kolumner och konstanter. Vanligtvis representerar kolumnerna variabler och varje rad representerar en observation. Då man startar `MINITAB` så får man ett nytt, tomt minnesutrymme. Så länge man är kvar i `MINITAB` finns de variabler man skapat kvar, men då man avslutar en session försvinner informationen. Om man har data som man vill använda igen kan man spara undan dessa i en fil (se kommandona **SAVE** och **RETRIEVE**). Denna fil kan man vid ett annat tillfälle sedan mata in i `MINITAB` igen.

Ett kommando består av själva kommandonamnet följt av noll, ett eller flera argument, eller subkommandon. Kommandonamnen är engelska ord. Man behöver inte skriva hela orden, utan för att `MINITAB` skall förstå räcker det med de fyra första bokstäverna i ordet. För att tala om för `MINITAB` att man vill ge några subkommandon, måste man avsluta kommandoraden med ett semikolon. `MINITAB` svarar med att ändra prompter till "SUBC >". Varje subkommando ges på en ny rad och avslutas med ett semikolon. Det sista subkommandot måste avslutas med en punkt. Om du glömmer att avsluta med ett semikolon eller en punkt så får du en varning av `MINITAB` . I så fall, skriv bara en punkt i sista raden och tryck på `<Return>` .

2.2 Help

Som de flesta programpaket har även `MINITAB` en inbyggd hjälpfunktion, som man når med ordet **help** följt av det man ville ha hjälp med. Detta är kanske det enda kommandot som man behöver kunna för att använda `MINITAB` . Anledningen till detta är att man kan få hjälp med alla kommandon när man kör programmet. För att se hur man använder kommandot skriver man bara **help help**. Om man skriver **help commands** så får man en lista över olika kategorier av kommandon som finns i `MINITAB` . Skriv t.ex. **help commands 13** och du får en lista över kommandon för statistisk kvalitetsstyrning.

3 Indata, utdata, editering och aritmetik

Här följer ett urval av de allra mest vanliga kommandona. Kommandonamnet är skrivet med stora bokstäver. Det som står inom hakparenteser behöver inte finnas med, och inte heller det som skrivs med små bokstäver. Bokstaven **C** står för kolumn. Alla kolumner måste ges namn på formen **C**n, där n är ett heltal. Konstanter ges på samma sätt namn enligt **K**n.[1]

3.1 Indata och utdata

- READ
  READ ['filnamn'] C1 ... Cn

  Om filnamnet utesluts så förutsätter `MINITAB` att man har för avsikt att mata in data från tangentbordet. Då man tryckt `<Return>` efter detta kommando skriver man helt enkelt in sina siffror rad för rad. Första raden kommer att innehålla den första observationens värden på de olika variablerna. Ett litet exempel:

  MTB > **READ C2 C3**
  DATA > **2 4**
  DATA > **3.5 27**
  DATA > **1 34**
  DATA > **END**
  3 ROWS READ

  Kolumn C2 kommer nu att innehålla värdena 2, 3.5 och 1. Efter att man talat om för `MINITAB` att man skall läsa in värden, byter `MINITAB` prompter till "DATA >" för att markera att det förväntas siffror.

- SET
  SET ['filnamn'] C
  Fungerar som **READ** med det undantaget att alla de värden man läser in hamnar i samma kolumn **C**. Tillsammans med **SET** kan man använda ett antal förkortningar. Antag t.ex. att man vill att kolumn C1 skall innehålla talen 1, 2, 3, 4, 5. Då skriver man helt enkelt:

[1] Riktigt sant är nu inte detta, vilket förstås av beskrivningen av kommandot **NAME**.

MTB > **SET C1**
DATA > **1:5**
DATA > **END**

Vill man ha en kolumn med fyra stycken ettor i, gör man på följande vis:

MTB > **SET C1**
DATA > **4(1)**
DATA > **END**

- INSERT
  INSERT ['filnamn'] [between K1 K2] C1 ... Cn
  Detta kommando används för att lägga till ett antal rader mellan rad K1 och K2 i kolumnerna **C1, ..., Cn**. Själva inläsandet av dessa rader fungerar på samma sätt som inläsning med **READ**.

- END
  Se exemplen i samband med **READ** och **SET**

- NAME
  NAME C1 = 'namn*1*' ... Cn = 'namn*n*'
  Om man vill slippa hålla reda på vilket kolumnnummer ens variabler, kolumner, har så kan man med detta kommandot ge dem namn. Dessa namn kan man sedan använda på alla ställen där `MINITAB` förväntar sig kolumnnummer.

- PRINT
  PRINT C1 ... Cn
  Skiver på skärmen ut vad man har i kolumnerna **C1, ..., Cn**.

3.2 Editering av data

- LET
  LET Ck(K1) = K2
  Med hjälp av detta kommando kan man ändra ett enda värde i en kolumn. I detta exemplet är det värdet i position **K1** i kolumnen **Ck** som får sitt värde ändrat till **K2**. De övriga förblir oförändrade.

- DELETE
  DELETE rows K1 ... Km of C1 ... Cn
  Det som står med små bokstäver här behöver inte tas med i kommandot, utan är med här bara som ett förtydligande. **DELETE** tar bort raderna **K1**, **K2** till och med **Km** ur kolumnerna **C1, ..., Cn**.

Dessutom finns det några andra kommandon som man kan läsa om i manualen eller titta på med hjälp av "**help** *commandname*" :

COPY	CODE	STACK
UNSTACK	CONVERT	CONCATENATE
ALPHA		

3.3 Aritmetik

LET E = expression
E kan antingen vara namnet på en konstant eller en kolumn. Några exempel:

MTB > **LET C1 = (C2 + C3) - 2**
MTB > **LET K1 = 45.78**
MTB > **LET C4 = C5 AND C6**

I det första exemplet kommer vart och ett av **C1**s element, **C1**i, att vara lika med **C2**i + **C3**i - 2. Det andra exemplet ger konstanten **K1** värdet 45.78. Operatorn **AND** i det sista exemplet är en logisk operator. Det innebär i det här fallet att **C4**i kommer att vara 1 om både **C5**i och **C6**i är större än 1 och annars 0.

Det finns givetvis fler operatorer och funktioner som kan finnas med i ett **LET**-uttryck. Nedan följer en sammanställning. Hur de fungerar kan man ta reda på med hjälp av `MINITAB` kommandot **help**.

RADVISA OPERATIONER (Dessa ger som resultat en ny kolumn)

Aritmetiska operatorer

+ addition	− subtraktion	* multiplikation
/ division	** upphöjt till	

Jämförelse operatorer

=	likhet	~=	inte lika med	<	mindre än
>	större än	<=	mindre än eller lika med	>=	större än eller lika med

Logiska operatorer

& eller **AND** | eller **OR** ~ eller **NOT**

Funktioner

ABSOLUTE	SQRT	LOGTEN	LOGE
EXPONENTIATE	ANTILOG	ROUND	SIN
COS	TAN	ASIN	ACOS
ATAN	SIGNS	PARSUMS	PARPRODUCTS
NSCORES			

KOLUMNVISA OPERATIONER

COUNT	N	NMISS	SUM
MEAN	STDEV	MEDIAN	MIN
MAX	SSQ	SORT	RANK
LAG			

Dessa finns även i radversion. Namnen börjar då med **R** och sedan följer kolumnmotsvarigheten.

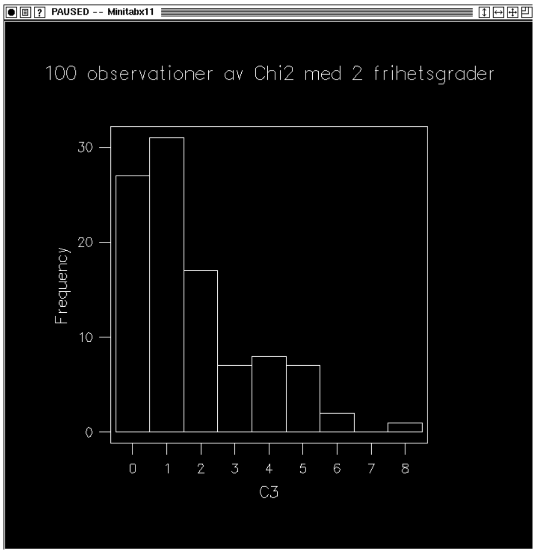
4 Diagram och plottar

4.1 Ett urval av grundkommandona

- HISTOGRAM

  HISTOGRAM C1 $\cdots$ **C**n

  Med hjälp av det här kommandot får man ett histogram för var och en av kolumnerna. Om du begärt histogram för mer än en kolumn kommer `MINITAB` bara att visa histogrammet för den första på skärmen. För att få se histogrammen för de nästkommande kolumnerna flyttar du *muspilen* till plot-fönstret och trycker på en valfri tangent. `MINITAB` visar nu den andra kolumnens histogram. Fortsätt på samma sätt för att se de övriga. Figure 4 visar hur det kan se ut när man gjort ett histogram med hjälp av `MINITAB`.



Figur 4: Histogramexempel

Om man bara ger kommandot **HISTOGRAM**, väljer `MINITAB` själv hur gruppindelningen skall göras. Genom att ge några subkommandon kan man själv betsämma sådana detaljer.

- TITLE = 'text'
- INCREMENT = K
- START = K1 [end = K2]
  Det första subkommandot genererar en titel till histogrammet. Med hjälp av de två andra styr man skalan på histogrammet. Det första reglerar avståndet mellan intervallens mittpunkter. Det andra anger mittpunkt för det första respektive sista intervallet.
- SAME
  Detta subkommando används då man skall göra fler än ett histogram och vill ha samma skala i alla.
- BY C
  Ett nytt histogram görs för vart och ett av de olika värdena i kolumen C, som skall vara av samma längd som de kolumner man vill ha histogram över. C måste dessutom innehålla heltalsvärden mellan -9999 och 9999. Kolumnen C fungerar alltså ungefär som en indikator på grupptillhörighet.

  De histogram som görs kommer alla att vara i samma skala.

Subkommmandona används på detta sätt:

MTB > **HISTOGRAM C1 C2 C3;** *Obs! semikolonet ingår*
SUBC > **SAME.** *Obs! punkten ingår*

- DOTPLOT
  DOTPLOT C1, ..., Cn
  Det här liknar ett histogram, men observationerna delas in i fler klasser. Egentligen en klass per observation, men för att alla observationer skall få plats på en horisontell linje måste `MINITAB` ibland gruppera ihop observationer.

  Till detta kommando finns precis samma subkommandon som till **HISTOGRAM** och de fungerar också på samma sätt.

- PLOT

  PLOT C1*C2

  `MINITAB` ritar en "scatterplot" med **C1** på y-axeln och **C2** på x-axeln.Varje punkt noteras med "$\diamond$".

  Till detta kommando finns en rad användbara subkommandon.

  - TITLE = 'text'

- FOOTNOTE = 'text'
- XLABEL = 'text'
- YLABEL = 'text'

Dessa fyra gör precis det man förväntar sig. Med hjälp av dem kan man alltså få en titel,en fotnot respektive namn på axlarna.

- SYMBOL = 'symbol'
  Om man ger detta subkommando kommer plotten att ritas med *symbol* istället för med $\diamond$.
- XINCREMENT = K
- XSTART = K1 [end K2]
- YINCREMENT = K
- YSTART = K1 [end = K2]
  Med hjälp av dessa fyra subkommandon kan man specificera skalan för x-axeln och y-axeln. **INCREMENT** styr avståndet mellan siffrorna längs med axlarna. **START** ger första och eventuellt sista punkt som skall plottas.

4.2 Att spara och skriva ut

Du kan spara plottar och histogram i grafiska filer. De här filerna kan du sedan använda i t.ex. laborationsrapporter.

Det finns ett subkommando till alla plot-kommandon som heter **GSAVE**. Om man anger ett filnamn tillsammans med subkommandot, kommer plotten att sparas i en fil med det namnet, med ett suffix som `MINITAB` lägger till;

MTB > **HISTOGRAM C1;**
SUBC> **GSAVE '*filnamn*'.**

För att kunna skriva ut måste man tala om för `MINITAB` vilket språk skrivaren använder. Man gör det med hjälp av kommandot **HARDCOPY** och det räcker att du gör det en gång per session. De flesta skrivare i Matematiskt Centrum använder postscript. Om du har tillgång till en sådan skrivare skriver du så här;

MTB > **HARDCOPY 'PSC'**

De filer du sparat med **GSAVE** kan du nu skriva ut med kommandot **GPRINT**;

MTB > **GPRINT '*filnamn*'**

eller titta på med kommandot **GVIEW**;

MTB > **GVIEW '*filnamn*'**

5 Något om några olika typer av filer i `MINITAB`

- SAVE
  SAVE ['filnamn']
  Efter att man använt detta kommando sparas allt man har i sitt minnesutrymme i filen *filnamn*.MTW. Användbart t.ex. då man vill sluta för dagen, men inte är färdig med bearbetningen av sitt datamaterial.

- RETRIEVE
  RETRIEVE ['filnamn']
  Detta kommando använder man då man vill läsa in en fil till `MINITAB` som man tidigare sparat undan med **SAVE**.

- OUTFILE
  OUTFILE ['filnamn']
  Med det här kommandot kan man spara sitt arbetspass i en fil. Allting som syns på skärmen sparas tills man skriver **NOOUTFILE**. Om man skriver **OUTFILE** igen med samma filnamn, läggs utdata till slutet av samma fil.

  Filen som sparas med **OUTFILE** har ASCII format och kan skrivas ut eller editeras med en editor. `MINITAB` lägger till LIS efter filnamnet.

- WRITE
  WRITE ['filnamn'] C1 ... Cn
  Om man angett ett filnamn, kommer innehållet i kolumnerna **C1 ... Cn** att skrivas ut till filen *filnamn*.DAT. Ändelsen .DAT lägger `MINITAB` själv till. Inne i `MINITAB` räcker det med *filnamn*, men utanför måste man ha med suffixet. Det är bara värdena i kolumnerna som skrivs ut, inte kolumnnamn eller nummer. Filen är en vanlig ASCII-fil som man kan skriva ut på vanligt sätt (**lpr** *filnamn*.DAT) från operativsystemet. Om man inte angett filnamn så dyker innehållet upp på skärmen istället.

6 Konfidensintervall och test

6.1 Konfidensintervall

- ZINTERVAL
  ZINTERVAL [K1% confidence], σ = K2 for C1, ..., Cm
  Här beräknas ett K1%-konfidensintervall för varje kolumn. De **m** kolumnerna antas vara **m** stycken stickprov från **m** olika normalfördelningar som alla har samma kända standardavvikelse, σ.

  Om man inte specificerat K1 så får man 95% intervall. Intervallen är på formen:

$$\{\ \overline{x} - z(\sigma/\sqrt{n}), \overline{x} + z(\sigma/\sqrt{n})\ \ \}$$

Här står $\overline{x}$ för medelvärdet av elementen i en kolumn, n är antalet element och z det värde från normalfördelningstabellen som motsvarar konfidensgraden. Förutom själva intervallet beräknar `MINITAB` också medelvärde, standardavvikelse och standardavvikelse / $\sqrt{n}$.

- TINTERVAL
  TINTERVAL [K% confidence] for C1, ..., Cm
  Ett separat t-konfidensintervall beräknas för var och en av de **m** stycken kolumnerna. Intervallen kommer att vara på formen:

  $$\{\ \overline{x} - t(s/\sqrt{n}), \overline{x} + t(s/\sqrt{n})\ \ \}$$

  $\overline{x}$ och $\sqrt{n}$ är som ovan. s står för standardavvikelsen och t är det värde i t-fördelningstabellen som motsvarar konfidensgraden.

6.2 Test

6.2.1 Parametriska test

- ZTEST
  ZTEST [μ = K1] σ = K2, for C1, ..., Cm
  Kommandot ser till att `MINITAB` utför ett dubbelsidigt z-test för var och en av kolumnerna. Om μ inte angivits kommer $\mu = 0$ att användas.

  Vill man istället få ett ensidigt test använder man subkommandot ALTERNATIVE K3. För K3 $= -1$ utförs testet mot alternativet $H_1 = \mu < K1$ och för K3 $= 1$ mot alternativet $H_1 = \mu > K1$, eller 0 istället för K1 om denna inte givits.

  `MINITAB`s utskrift efter kommandot består av medelvärdet, standardavvikelsen, medelvärdets standardavvikelse, värdet på test-statistikan Z och p-värdet. Här är:

  $$Z = (\overline{x} - \mu)/(\sigma/\sqrt{n})$$

- TTEST
  TTEST [μ = K1] on C1, ..., Cm
  Utför ett t-test per kolumn. Det fungerar precis på samma sätt som för ZTEST ovan. Skillnaden är att här beräknas test-statistikan

  $$T = (\overline{x} - \mu)/(s/\sqrt{n})$$

  istället.

- TWOSAMPLE
  TWOSAMPLE [K1% confidence] for C1 C2
  Med hjälp av detta kommndo utförs ett två-stickprovs t-test och beräknas ett konfidensintervall. Låt $\mu_1, \overline{x}_1, s_1$ och n_1 vara den första populationens väntevärde, stickprovs medelvärde, stickprovs standardavvikelse samt stickprovsstorlek. Låt på samma sätt $\mu_2, \overline{x}_2, s_2$ och n_2 vara det samma för den andra populationen. De två stickproven antas vara oberoende. Kommandot beräknar ett K1% konfidensintervall för $\mu_1 - \mu_2$:

  $$\{ \ \ (\overline{x}_1 - \overline{x}_2) - ts, (\overline{x}_1 - \overline{x}_2) + ts) \ \}$$

  och utför testet $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 \neq \mu_2$.

  Testet baseras på test-statistikan:

  $$T = (\overline{x}_1 - \overline{x}_2)/s$$

  där $s = \sqrt{{s_1}^2/n_1 + {s_2}^2/n_2}$. De två populationerna antas inte ha samma varians.

  Kommandot har två subkommandon. Dels ALTERNATIVE som fungerar som för de andra två testkommandona, dels POOLED. Då man ger subkommandot POOLED kommer `MINITAB` att istället för s ovan använda den sammanvägda stickprovs standardavvikelsen.

- TWOT
  TWOT [K% confidence] data in C1, groups in C2
  Detta kommando utför precis samma saker som kommandot ovan. Skillnaden ligger i hur data anges. TWOSAMPLE förväntar sig data från det första stickprovet i den första kolumnen och data från det andra i den andra. TWOT däremot vill ha alla data i den första kolumnen. Den andra kolumnen skall specificera vilken grupp ett värde i den första tillhör. Kolumn C2 fungerar alltså ungefär som C i BY C, som figurerat som subkommando bl.a. till HISTOGRAM.

- CHISQUARE
  CHISQUARE test on table stored in C1, ..., Cn
  Utför ett χ^2-test på den kontingenstabell som finns sparad i kolumnerna **C1** till och med **Cn**. Om du inte har dina data i tabellform, så kan du använda kommandot **TABLE** tillsammans med subkommandot **CHISQUARE**. Se med hjälp av **help** hur det fungerar.

  Här kommer ett litet exempel på hur **CHISQUARE** fungerar:

  MTB > **READ C1-C3**
  DATA > **28 18 4**
  DATA > **22 27 1**
  DATA > **END**
  MTB > **CHISQUARE C1-C3**

Expected counts are printed below observed counts

	C1	C2	C3	Total
1	28	18	4	50
	25.00	22.50	2.50	
2	22	27	1	50
	25.00	22.50	2.50	
Total	50	45	5	100
Chisq =	0.36 +	0.90 +	0.90 +	
	0.36 +	0.90 +	0.90 =	4.32

df = 2
2 cells with less expected counts then 5.0

6.2.2 Ickeparametriska test

- STEST
  STEST [median = K] on C1, ..., Cn
  Utför ett teckentest av $H_0 : median = K$ mot $H_1 : median \neq K$ för var och en av kolumnerna **C1** till och med **Cn**. Om medianen inte specificerats så används värdet 0.

- WTEST
  WTEST [of median = K] on C1, ..., Cn
  Ett enstickprovs Wilcoxon-test av $H_0 : median = K$ mot $H_1 : median \neq K$ utförs för kolumnerna **C1** till och med **Cn**. På samma sätt som för kommandot ovan kommer $K = 0$ att användas om inget annat värde angivits.

- MANN-WITNEY
  MANN-WITNEY [confidence = K] on C1 C2
  Med detta kommando kommer ett tvåstickprovs rang-test att göras på data i kolumnerna **C1** och **C2**. Det test som utförs kallas ofta Mann-Witney eller två-stickprovs Wilcoxon-test.

  Låt η_1 vara medianen i den första populationen, från vilken det finns ett stickprov i kolumn **C1**, och η_2 vara den andra populationens median. Det test som utförs är alltså $H_0 : \eta_1 = \eta_2$ mot $H_1 : \eta_1 \neq \eta_2$. Liksom sina parametriska motsvarigheter, TWOSAMPLE och TWOT, beräknar Mann-Witney-kommandot även ett konfidensintervall.

Precis som för de parametriska testkommandona finns subkommandot **ALTERNATIVE**. Om man ger **ALTERNATIVE -1** kommer mothypotesen att bli $H_1 : \eta_1 < \eta_2$ och om man ger **ALTERNATIVE 1** blir mothypotesen $H_1 : \eta_1 > \eta_2$.

7 Variansanalys

Det finns två kommandon för ensidig variansanalys.

- AOVONEWAY
  AOVONEWAY on C1 ... Cn
  Med hjälp av detta kommando utförs en ensidig variansanalys. Olika kolumner representerar observationer från olika grupper.

- ONEWAY
  ONEWAY aov, data C, levels in C
  Det här är samma som AOVONEWAY med skillnaden att man har alla observationer i en kolumn. Den andra kolumnen specifiserar vilken grupp ett värde i den första tillhör. Om man inte har sina data på det sättet kan man lätt överföra det dit.

  Antag att vi har data för 3 grupper i kolumnerna $C1$, $C2$ och $C3$. Vi gör nu på följande sätt;

  MTB > **STACK C1, C2, C3, C4;**
  SUBC > **SUBSCRIPTS C5.**

  De kolumner som skall ges till **ONEWAY** är data-kolumnen **C4** och gruppvektorn **C5**

7.1 Bonferroni- och Tukey-intervall

Det finns inget kommando eller subkommando som direkt ger Bonferroniintervallen. Däremot kan man få tabellvärden till de konfidensgrader man själv räknat fram;

MTB > **INVCDF 0.9875;**
SUBC > **t 4.**

På detta sätt får vi reda på det x som uppfyller $P(X \leq x) = 0.9875$, där X är en stokastisk variabel som är t-fördelad med frihetsgradstalet 4.

Tukey-intervall kan man få genom att ge ett subkommando till ANOVA-kommandot **ONEWAY**. Antag att vi har data i kolumnen **C1** och deras respektive grupptillhörighet i kolumnen **C2**. För att då få Tukey-intervall med den simultana konfidensgraden **K** gör man så här;

MTB > **ONEWAY C1 C2;**
SUBC > **TUKEY K.**

7.2 Kruskal-Wallis

Det finns ett antal ickeparametriska alternativ till variansalysen. Här presenteras ett, nämligen Kruskal-Wallis testet. Nollhypotesen att det inte finns någon skillnad mellan de olika populationerna, testas mot att det finns någon skillnad.

- KRUSKAL-WALLIS
  KRUSKAL-WALLIS test for data in C1, levels in C2
  Värdena i den andra kolumnen talar om vilken population motsvarande värde i den första kolumnen tillhör.

8 Regressionsanalys

REGRESS C1 on K pred C2,...,Ck

Kommandot **REGRESS** använder minsta kvadratmetoden för att anpassa en modell till en eller flera bakgrundsvariabler. Modellen som anpassas då **K**=1 är;

$$Y = \beta_0 + \beta_1 X + e$$

Här betecknar Y responsvariabeln, X prediktorn samt β_0 och β_1 regressionskoefficienterna. e är en normalfördelad felterm med väntevärde 0 och standardavvikelse σ. Då n st observationer använts skattar **REGRESS** på vanligt sätt enligt:

β_0 med $b_0 = \bar{y} - b_1\bar{x}$

β_1 med $b_1 = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^{n}(x_i - \bar{x})^2$

σ med $s = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-2)}$

Den anpassade ekvationen blir nu:

$$\hat{Y} = b_0 + b_1 X$$

$\hat{Y}$ kallas prediktionen eller det anpassade värdet.

8.1 Utdata

Om man bara ger regressionskommandot får man bl.a. följande utdata:

- Regressionslinjen.

- En tabell över skattningarna av koefficienterna. Tabellen innehåller dessutom skattningarnas standardavvikelse, t-värde vid test om de är lika med 0 samt p-värdet vid detta test.

- s, skattningen för σ.

- R-sq, R^2, även kallad förklaringsgraden.

- En tabell över extrema observationer.

Om man dessutom vill få en tabell över anpassade värden, deras standardavvikelser samt residualer och deras standardavvikelser, så får man ge ett speciellt kommando till `MINITAB` innan man utför själva regressionen;

MTB > **BRIEF 3**
MTB > **REGRESS ...**

Kommandot **BRIEF** styr hur mycket som **REGRESS** lämnar ifrån sig. Följande kommer att hända för de olika värden man kan ge **BRIEF**;

- **K = 0:** Ingen utskrift alls.

- **K = 1:** Regressionsekvationen, tabell över koefficienterna, s, R^2, anpassade R^2 och första delen av en variansanalystabell.

- **K = 2:** Samma utskrift som om man inte använt **BRIEF**, d.v.s. förutom det man får om man använder **1** dessutom resten av anova-tabellen samt en tabell över extrema värden.

- **K = 3:** Förutom det som skrivs ut vid **K = 2**, dessutom en tabell över anpassade värden och residualer.

8.2 Prediktion

Man kan prediktera responsvariabeln för olika värden på bakgrundsvariablerna. Det gör man med hjälp av subkommandot:
PREDICT for E2,...,Ek

Som resultat får man en tabell som innehåller de anpassade värdena på responsvariabeln, standardavvikelsen för dem, ett 95% konfidensintervall och ett 95% prediktionsintervall.

9 Simulering av stokastiska variabler

För att generera slumptal från olika fördelningar finns kommandot **RANDOM**. Syntaxen är som följer:
RANDOM K observations into each of C1, ..., Cn
Om det bara ges så här, utan något subkommando, kommer **n** st stickprov innehållande vardera **K** observationer av $N(0, 1)$-variabler att simuleras. Resultatet läggs i kolumnerna **C1**, ..., **Cn**. Med hjälp av subkommandon kan man styra vilken fördelning som skall simuleras. Om man t.ex. vill ha två stickprov om tio observationer från en Poisson-fördelning med $\mu = 0.5$ skriver man:

MTB > **RANDOM 10 C1 C2;**
SUBC > **POISSON 0.5.**

I detta sammanhang kan även några andra kommandon nämnas. De har samma subkommandon som RANDOM.

- PDF
  PDF for values in E1 [put into E2]
  `MINITAB` beräknar värdet för täthetsfunktionen för de värden som finns angivna i **E1**.

- CDF
  CDF for values in E1 [put into E2]
  `MINITAB` beräknar värdet för fördelningsfunktionen för de värden som ges i **E1**.

- INVCDF
  INVCDF for values in E1 [put into E2]
  Inversen till fördelningsfunktionen kommer att beräknas, d.v.s för en given sannolikhet p kommer **INVCDF** att returnera det värde x som upfyller P($X \leq x$) = p. Här står X för en stokastisk variabel med fördelningen N(0,1) eller given med hjälp av subkommando.

9.1 Subkommandon till RANDOM

- BERNOULLI
  BERNOULLI p = K
  Bernoulli-fördelning. p är sannolikheten för att lyckas.

- BINOMIAL
  BINOMIAL n = K1 p = K2
  Binomial-fördelning. n är antalet försök och p sannolikheten att lyckas i vart och ett av försöken.

- POISSON
  POISSON μ = K
  Posson-fördelning. Värdet på μ måste ligga mellan 0 och 50.

- NORMAL
  NORMAL [μ = K1 [σ = K2]]
  Normal-fördelning med väntevärde μ och standardavvikelse σ. Om något av K:na utesluts så använder `MINITAB` $\mu = 0$ respektive $\sigma = 1$.

- UNIFORM
  UNIFORM [a = K1 b = K2]
  Likformig-fördelning på intervallet från a till b. Om inga parametrar anges kommer `MINITAB` att använda $a = 0$ och $b = 1$.

- CAUCHY
  CAUCHY [a = K1 [b = K2]]
  Cauchy-fördelning. Om a och b inte specificeras använder `MINITAB` 0 respektive 1. Täthetsfunktionen ser ut så här:

$$\mathrm{f}(x) = 1/\{(\pi b)[1 + \{(x-a)/b\}^2]\}$$

- EXPONENTIAL
  EXPONENTIAL [b = K]
  Exponential-fördelning. Om b inte specifiserats används $b = 1$. Täthetsfunktionen ser ut så här:

$$\mathrm{f}(x) = \frac{1}{b}e^{-x/b} \qquad x > 0$$

- WEIBULL
  WEIBULL a = K1 b = K2
  Weibull-fördelningen. Täthetsfunktionens utseende:

$$\mathrm{f}(x) = (ax^{a-1}e^{-(x/b)^a})/b^a \qquad x > 0$$

Det finns ytterligare några fördelningar att tillgå . Titta på dem med hjälp av **help** om ni är intresserade.

INTEGER	T	F	LAPLACE
DISCRETE	CHISQUARE	LOGNORMAL	LOGISTIC
GAMMA	BETA		

10 Statistisk kvalitetsstyrning

MINITAB har elva kommandon för att rita olika kontrolldiagram. Vi börjar med en kort genomgång av några subkommandon som kan användas i kombination med *alla* kontrolldiagram och sedan går vi genom kommandot för $\bar{X}$-diagram.

- SLIMITS
  SLIMITS are K1 ... Kn
  Det här subkommandot definierar kontrollgränserna i diagrammet. T ex om man skriver **SLIMITS 1,2**, plottas fyra kontrollgränser i diagrammet, två linjer i varje sida av centrallinjen med avståndet 1σ och 2σ från centrallinjen.

- GSAVE
  GSAVE 'filename'
  Man använder det här subkommandot för att spara undan ett diagram i en fil. I avsnitt 4.2 finns beskrivet hur man använder det.

  Följande subkommando används för att skriva text i olika delar av diagrammet (se kommandot **PLOT**).

- **XLABEL 'text'**
- **YLABEL 'text'**
- **TITLE 'text'**
- **FOOTNOTE 'text'**

10.1 $\bar{X}$-diagram

För att plotta $\bar{X}$-diagram för olika variabler måste man ha olika observationer av varje variabel i en kolumn. Syntaxen är som följer:

XBARCHART for C subgroups are in E

Det här kommandot ritar ett $\bar{X}$-diagram för kolumnen **C**. Kolumnen **C** antas innehålla ett antal stickprov från en variabel. Antalet observationer i varje stickprov definieras med **E**. Om t ex är **E** lika med 5, så tar MINITAB första 5 raderna i kolumnen som första stickprovet på variabeln, andra 5 observationerna för andra stickprovet osv. MINITAB antar att alla observationer kommer från en normalfördelning och beräknar medelvärdet och stickprovets standardavvikelse. Om man vill, kan man ange väntevärdet och standardavvikelsen med de här subkommandona:

- **MU = K**
- **SIGMA = K**

I så fall antar `MINITAB` att observationerna kommer från normalfördelning med väntevärdet **MU** och standardavvikelsen **SIGMA**.

10.2 Andra kontrolldiagram

I detta sammanhang kan vi nämna följande kommando som används för att rita andra typer av kontrolldiagram i `MINITAB` .

- **RCHART** ritar ett R-diagram
- **SCHART** ritar ett s-diagram
- **%XRCHART** ritar ett $\bar{X}$-diagram och ett R-diagram
- **%XSCHART** ritar ett $\bar{X}$-diagram och ett S-diagram
- **MACHART** ritar ett glidande medelvärdesdiagram
- **EWMACHART** ritar ett exponentiellt viktat diagram
- **PCHART** ritar ett p-diagram
- **NPCHART** ritar ett np-diagram
- **CCHART** ritar ett c-diagram
- **UCHART** ritar ett u-diagram

Skriv "**help** *commandname*" för att se hur man använder kommandona och vilka subkommandon finns för varje diagram.

11 Diverse mer eller mindre "statistiska" kommandon

- INFO
  INFO
  Det här kommandot ger en kort information om alla kolumner, med deras eventuella namn, och konstanter som finns i minnesutrymmet.

- DESCRIBE
  DESCRIBE C
  Om man vill få en kort beskrivning av sin variabel, kolumn, så kan man använda kommandot DESCRIBE. Det förklaras lättast med ett litet exempel:

  MTB > **SET C1**
  DATA > **10, 13, 9, 15, 17, 20, 15, 17, 8, 11, *, 12, 14**

DATA > **END**
MTB > **DESCRIBE C1**

	N	N*	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
C1	12	1	13.42	13.50	13.30	3.60	1.04
	MIN	MAX	Q1	Q3			
C1	8.00	20.00	10.25	16.50			

N	Antalet värden i kolumen	SEMEAN	STDEV / $\sqrt{N}$
N*	Antalet "missing value"	MIN	Minsta värdet i kolumnen
MEAN	Medelvärdet av talen i kolumnen	MAX	Största värdet i kolumnen
TRMEAN	Det trimmade medelvärdet (d.v.s. då man tagit bort de 5% minsta och 5 % största värdena)	Q1	25% percentilen
		Q3	75% percentilen
STDEV	Standardavvikelsen		

- COVARIANCE
  COVARIANCE for C1, ..., Cm [put into M]
  Detta kommando ser till att `MINITAB` beräknar kovariansen mellan varje par av kolumnerna som angivits. Om det finns "missing-values", tas bara de rader där det finns värden i båda kolumnerna med i beräkningarna. Om man sparar undan kovarianserna i matrisen M, skrivs de inte ut på skärmen.

- CORRELATION
  CORRELATION for C1, ..., Cm [put into M]
  Här kommer korrelationen mellan varje par av kolumnerna att beräknas. Kommandot fungerar på samma sätt som COVARIANCE.

- CENTER
  CENTER C1, ..., Cm put into C(m+1), ..., C(m+m)
  Med hjälp av detta kommando kan man standardisera värdena i sina kolumner. Använd **help** för att se vilka subkommandon som finns.

- SORT
  SORT C1 put in C2
  Sorterar värdena i kolumnen **C1** och lägger resultatet i kolumnen **C2**. Det finns varianter på detta kommando, vilka man kan studera med hjälp av **help sort**.

- RANK
  RANK the values in C1, put ranks into C2
  I kolumnen **C2** kommer rangerna för motsvarande värden i kolumnen **C1** att läggas.

- NSCORES
  NSCORES of C1, put into C2

För att göra en ”normal probability plot” måste detta kommando kombineras med kommandot **PLOT** på följande sätt:

MTB > **NSCORES C1 C2**
MTB > **PLOT C1*C2**

Referenser

[1] MINITAB Reference Manual - Release 9, december 1992.