



EKONOMIHÖGSKOLAN

Lunds universitet

Statistiska institutionen

LWn/PEI/JB

Introduktion till Minitab version 14

Innehållsförteckning

1	Introduktion	sid 2
	Worksheeten – datafönstret	
	Minitabs menyer och Session-fönstret	
	Att spara och öppna Minitab-filer	
	”Missing values” i Minitab	
2	Att skapa nya variabler – transformationer	sid 7
	Calc > Calculator	
	Data > Code	
3	Beskrivande statistik	sid 9
	Frekvenstabeller, korstabeller	
	Medelvärden o sånt	
4	Några vanliga diagramtyper	sid 15
	Stapeldiagram, histogram, boxplot	
	Layout-verktyget	
5	Hur man kopierar resultaten till Word	sid 23
6	Konfidensintervall och hypotesprövning	sid 24
	Parvisa observationer, två oberoende stickprov	
	Procenttal	

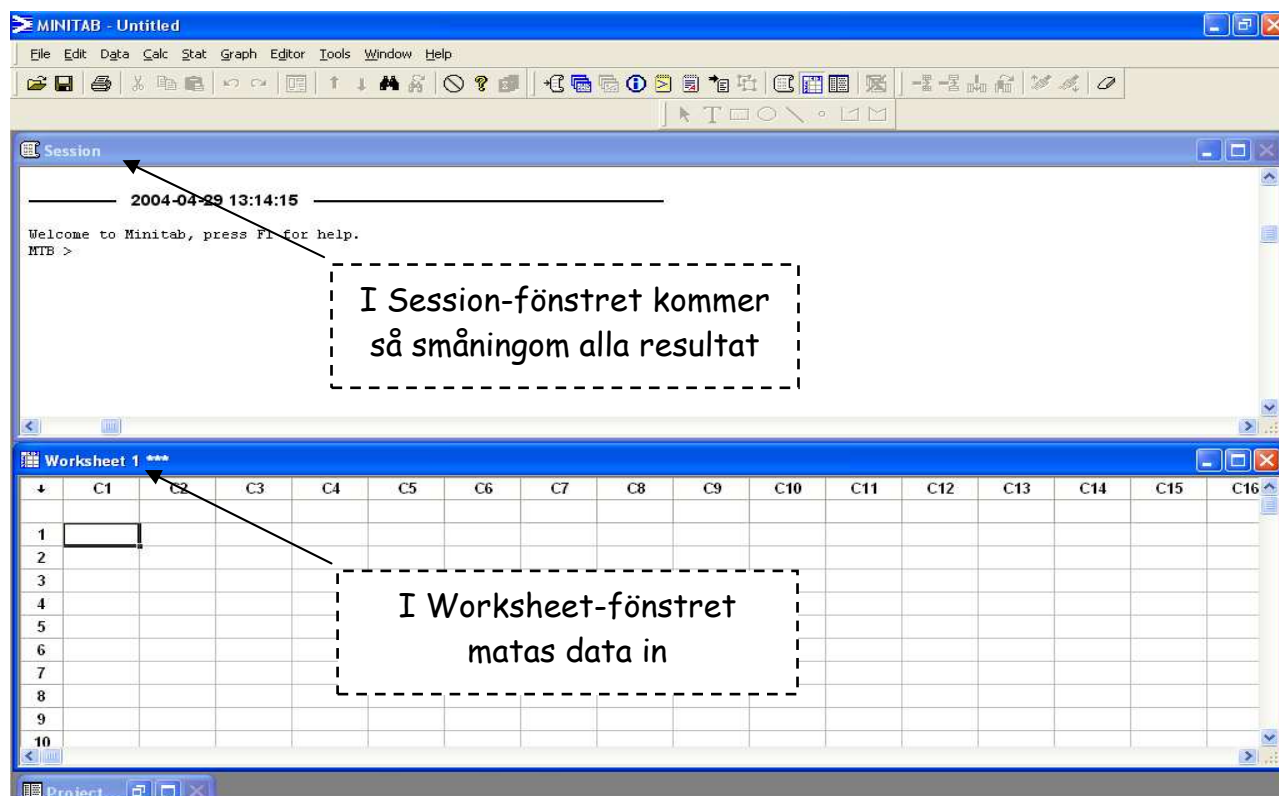
1. Introduktion

Datorprogram för bearbetning av statistiska data har funnits nästan lika länge som datorer har funnits. På senare år har även kalkylprogram som Excel, Lotus m fl fått utökade möjligheter men när det gäller statistisk analys är de riktiga statistikprogrammen utan tvekan bäst lämpade.

Minitab från Minitab Inc., Pennsylvania, USA är ett av de mest använda och spridda statistikprogrammen över hela världen. Programmet utvecklades ursprungligen vid Pennsylvania State University för användning vid universitetskurser och spreds snabbt över hela universitetsvärlden tack vare sitt enkla och effektiva koncept. Verksamheten avknoppades och är idag ett kommersiellt företag med ca 200 anställda varav en del finns i Europa. Mer information hittar du på www.minitab.com. Där finns det också möjlighet att gratis ladda ner en demoversion som fungerar i 30 dagar.

1.1 Worksheeten – datafönstret

Så här ser det ut i skärmen när man startat Minitab.



Två fönster dyker upp. Det översta är **Session-fönstret** där alla resultat – utom grafiska bilder – hamnar och det understa är en **Worksheet** – det rutnät där alla data matas in. Varje diagram som ritas kommer i ett eget fönster.

När data (information) lagras i de olika kolumnerna (C1, C2, ...) i worksheeten måste man komma ihåg att ...

Vid statistiska bearbetningar matas data alltid in så att varje rad motsvarar en individ och varje kolumn motsvarar en variabel (t ex ålder, kön, etc).

Vi har i en grupp studenter gjort en enkät med bland annat dessa frågor.

1. Namn: _____

2. Kön: Man (1) Kvinna (2)

3. Födelseår: 19____

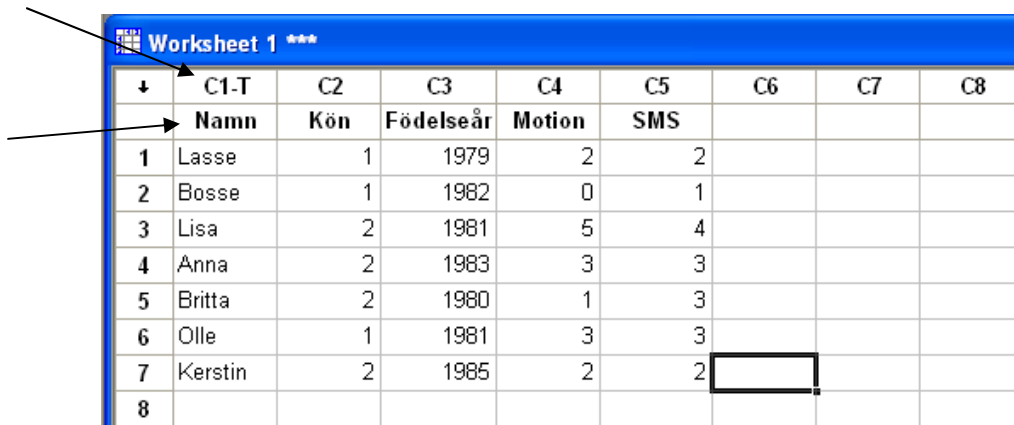
4. Hur många gånger per vecka motionerar du?: _____ gånger

5. Hur ofta skickar du SMS?:

Aldrig (1) Någon gång/vecka (2) Flera gånger/vecka (3) Dagligen (4)

Svaren på de fem frågorna ska matas in i fem kolumner. Men innan svaren matas in i Minitabs worksheet – ”rutnätet” – är det praktiskt att koda svaren på frågorna 2 och 5 till siffervärden. Inmatningen går mycket snabbare då. På fråga 2 kodar vi svaret Man till 1 (ett) och Kvinna till 2. På den femte frågan låter vi 1 = Aldrig, 2 = Någon gång, 3 = Flera gånger och 4 = Dagligen.

När de sju första frågeformulären är inmatade ser det ut så här i bildskärmen.



	C1-T	C2	C3	C4	C5	C6	C7	C8
	Namn	Kön	Födelseår	Motion	SMS			
1	Lasse	1	1979	2	2			
2	Bosse	1	1982	0	1			
3	Lisa	2	1981	5	4			
4	Anna	2	1983	3	3			
5	Britta	2	1980	1	3			
6	Olle	1	1981	3	3			
7	Kerstin	2	1985	2	2			
8								

På den första raden hittar vi Lasse som är Man. Han är född 1979, motionerar två gånger per vecka och skickar SMS någon gång per vecka. På de följande raderna följer sedan resten av formulären. Ytterligare två saker ska vi också lägga märke till (pilarna i bilden):

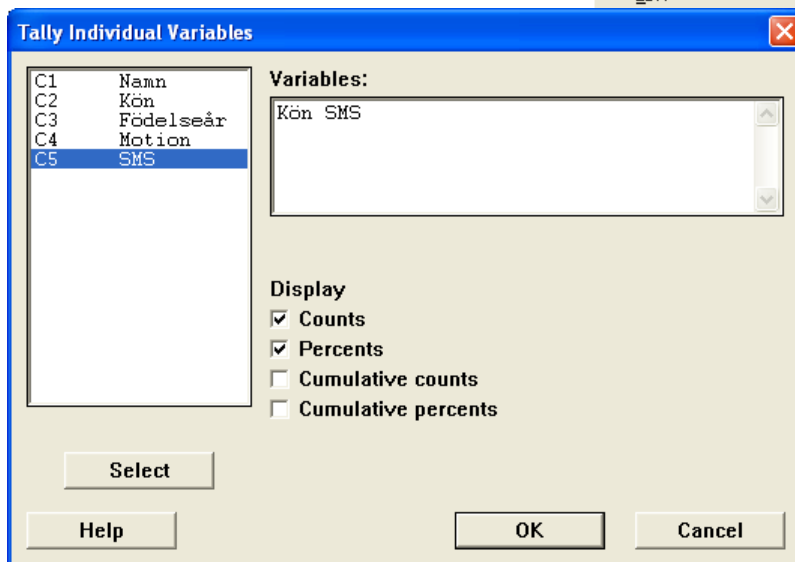
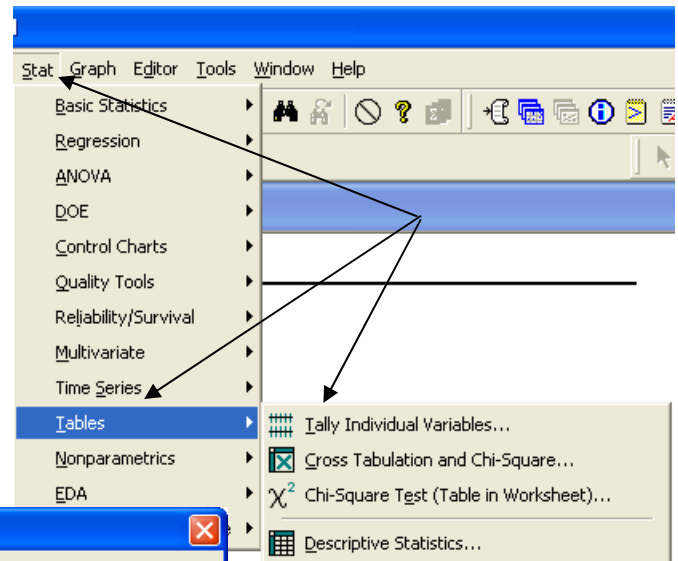
- I samma ögonblick vi matade in namnet Lasse överst i kolumn C1 fick den -T efter sig. Det innebär att kolumnen innehåller text. De övriga kolumnerna innehåller numeriska data. Minitab kan även hantera datum och klockslag.
- Vi har också passat på att ge kolumnerna namn som gör det lättare för oss att hålla reda på vad de innehåller; Namn, Kön, Födelseår, Motion och SMS.

1.2 Minitabs menyer och Session-fönstret

Att göra en analys på sju observationer är kanske inte så meningsfullt. Här gör vi det bara för att visa hur resultaten presenteras i Session-fönstret.

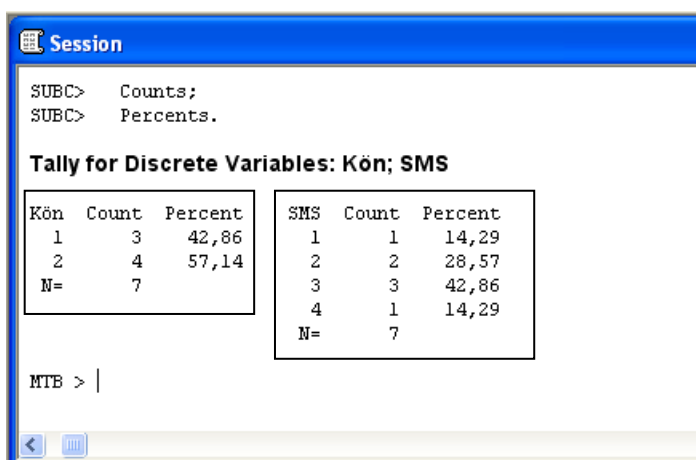
Vi vill ha frekvenstabeller för variablerna kön (C2) och SMS (C5). Man börjar med att på den översta menyraden (se bilden till höger) klicka på ordet **Stat** (= Statistisk analys). Därefter väljs **Tables** och slutligen **Tally Individual Variables**.

Stat > Tables > Tally ... skrivs detta i häftet i fortsättningen.



När man klickat på dessa alternativ får man upp menyn till vänster i skärmen.

Där markerar man i tur och ordning de båda variablerna i den stora rutan till vänster och när man klickar på **Select** åker de över till den högra rutan. Alternativt kan man *dubbeltlicka* på dem så åker de över direkt.



Avsluta med ett klick på OK och resultatet dyker upp i Session-fönstret som du ser här till vänster.

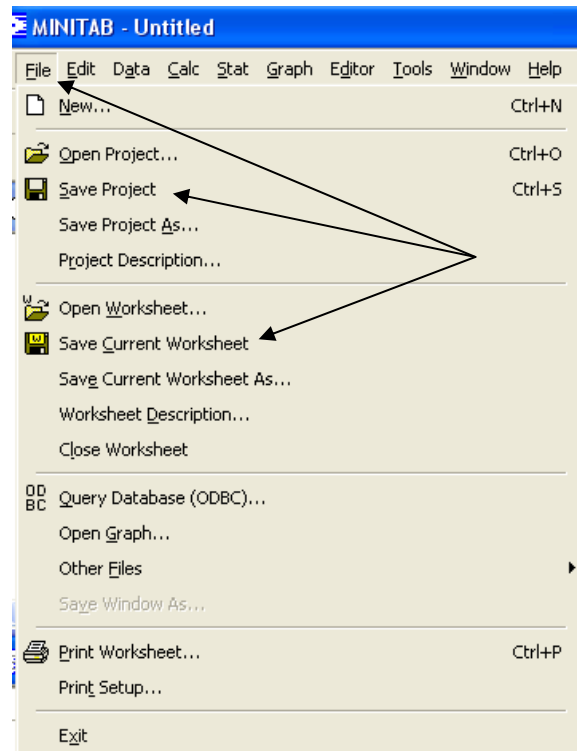
Två enkla frekvenstabeller har skapats. Vi ser t.ex. könsfördelningen tre män (1:or) och fyra kvinnor (2:or), 42,86 % mot 57,14 %. (Ramarna runt tabellerna är ditlagda efteråt!)

Men mer om tabeller och annat statistiskt längre fram. Bland annat ska vi se hur man kan förvandla 1:or till texten Man och 2:or till Kvinna.

1.3 Att spara och öppna Minitab-datafiler

Det finns två sätt att spara data:

- **File > Save Project (As)** eller **Ctrl+S** innebär att alla data i worksheeten (ja man kan t.o.m. ha flera worksheets öppna samtidigt) plus hela innehållet i Session-fönstret samt alla bilder man skapat och ytterligare en del information sparas som en enda *Minitab Project*-fil. Dessa filer får filändelsen *.MPJ*.
- **File > Save Current Worksheet (As)** innebär att man enbart sparar de data som finns i worksheeten. Eftersom man ofta flyttar över alla viktiga resultat från Session-fönstret och grafikfönster till ett ordbehandlingsprogram finns som regel ingen anledning att spara det även som ett Minitab-projekt. Filändelsen (extension) *.MTW* får filerna automatiskt när man väljer detta alternativ.
- Worksheeten kan också enkelt sparas i format som kan läsas av andra program, t.ex. Excel, dBase, Quattro Pro. Välj **Save Worksheet** och *ändra filformatet* till det som önskas.



Sitter du i statistiska institutionens PC-rum och jobbar kan du endast spara dina filer på Mina dokument: eller den enhet som heter H:. Datorerna är kopplade till en server och saknar egen hårddisk.

På enheterna M: och S: ska du ibland hämta filer till de olika uppgifterna men du kan inte spara på dessa enheter.

Filer öppnas på motsvarande sätt som de sparades.

- **File > Open Worksheet** öppnar en sparad worksheet. Kom ihåg att en worksheet bara innehåller de data som fanns i "rutnätet" när den sparades. Inga resultat från Session-fönstret eller bilder finns här. Dessa har du säkert redan flyttat över till Word och sparat som en del av din rapport.
- **File > Open Project** eller **Ctrl+O** öppnar en sparad projektfil. Alla data, bilder och allt i Session-fönstret kommer fram.
- Excel-, dBase-filer m.fl. kan också öppnas direkt i Minitab. **File > Open Worksheet** och ändra filformatet till Excel, dBase eller vad du nu har för fil. Kom bara ihåg att den fil du försöker öppna måste ha data på det "statistiskt korrekta sättet", dvs varje rad är en individ och varje kolumn motsvarar en variabel. På översta raden kan man ha variabelnamnen.

1.4 ”Missing values” i Minitab

Allt för ofta råkar man ut för att inte alla personer besvarar alla frågor eller att det av andra skäl saknas några datavärden. I Minitab används en asterisk (*) som missing value-kod för numeriska data. Gäller det tecken-data, som förnamnet, kan man bara låta rutan vara tom om man vill.

Antag att anonyma Anna i vår enkät inte ville uppge sitt förnamn. Dessutom skämdes Bosse över sina dåliga motionsvanor så han besvarade varken den frågan eller frågan om kön. Då skulle vår worksheet ha sett ut så här i stället.

↓	C1-T	C2	C3	C4	C5	C6	C7	C8
	Namn	Kön	Födelseår	Motion	SMS			
1	Lasse	1	1979	2	2			
2	Bosse	*	1982	*	1			
3	Lisa	2	1981	5	4			
4		2	1982	3	3			
5	Britta	2	1980	1	3			
6	Olle	1	1981	3	3			
7	Kerstin	2	1985	2	2			
8								

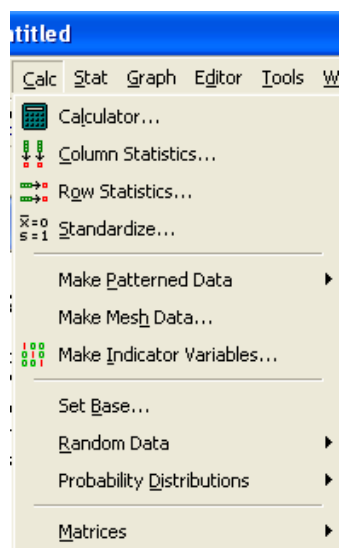
Blankt i namnkolumnen och asterisker i de numeriska kolumnerna.

2 Att skapa nya variabler – transformationer

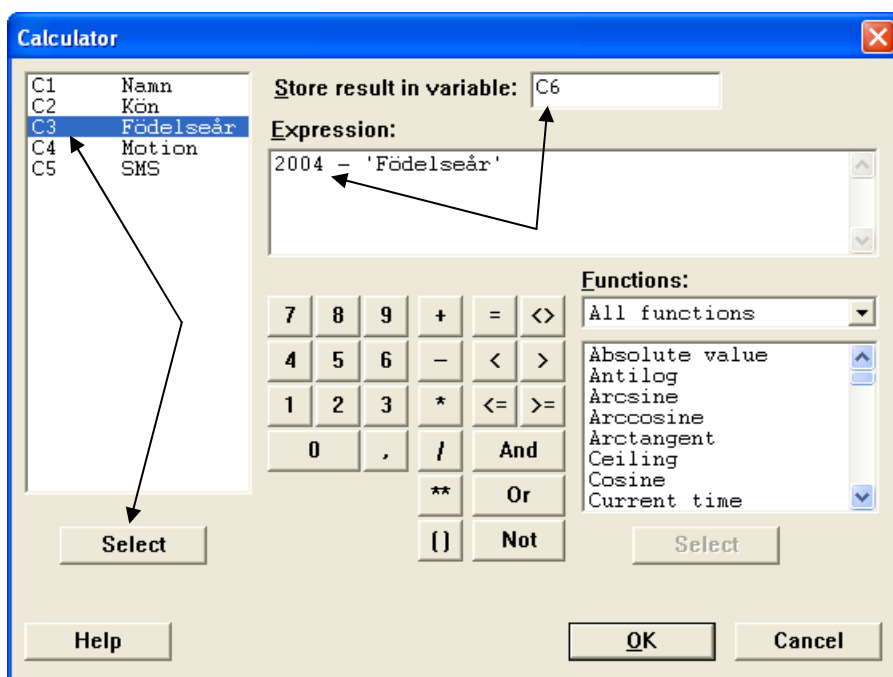
Det finns många tillfällen när man behöver göra beräkningar eller ”manipulera” data olika sätt. Vi kodade kön till 1/2 men vill hellre ha Man/Kvinna i resultattabellerna. Vi frågade inte efter personernas ålder men kan ju med hjälp av födelseåret räkna ut hur gamla personerna är. När vi sedan räknat ut personernas ålder vill vi kanske gruppera dem i olika åldersgrupper – Unga/Medelålders/Gamla.

2.1 Calc > Calculator

Calc > Calculator används när den nya variabeln skapas med ett matematiskt uttryck. Vi kan beräkna åldern vid årets slut 2004 för personerna i vår enkät genom att subtrahera födelseåret från 2004; Ålder = 2004 – Födelseår. Så här görs det med Minitab.



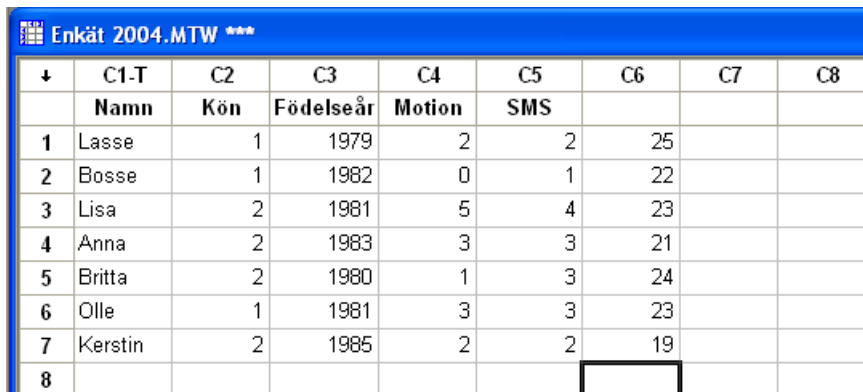
Klicka på **Calc** på menyraden överst i skärmen. Till vänster ser vi de möjligheter som finns där. Välj sedan **Calculator** och då kommer nedanstående meny fram. Första lediga kolumn är C6 så det skriver vi i rutan för resultatvariabeln.



I ”Expression:”-rutan skriver man in 2004–C3. Det matematiska uttryck som beräknar åldern.

Alternativt kan man klicka in uttrycket och när man då väljer C3 i den vänstra rutan och klickar på Select så kommer namnet 'Födelseår' att synas i stället.

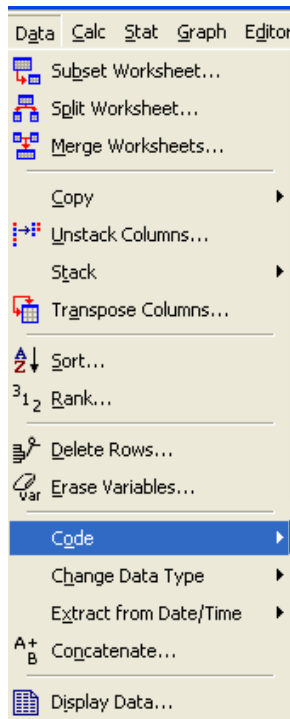
Resultatet ser vi här intill. Det enda som återstår är att sätta namn på kolumn C6.



	C1-T	C2	C3	C4	C5	C6	C7	C8
	Namn	Kön	Födelseår	Motion	SMS			
1	Lasse	1	1979	2	2	25		
2	Bosse	1	1982	0	1	22		
3	Lisa	2	1981	5	4	23		
4	Anna	2	1983	3	3	21		
5	Britta	2	1980	1	3	24		
6	Olle	1	1981	3	3	23		
7	Kerstin	2	1985	2	2	19		
8								

2.2 Data > Code

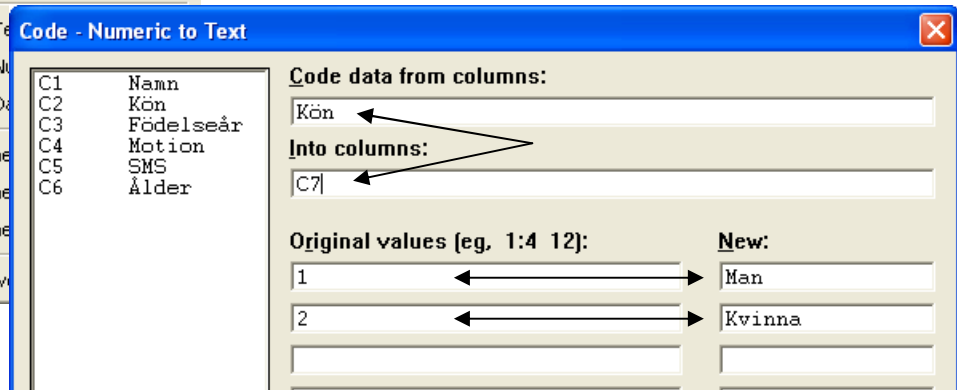
Problemet med Man/Kvinna i stället för 1/2 löser man enkelt under **Data > Code**. Bland alla omkodningsalternativ väljer vi **Numeric to Text** – vår numeriska variabel ska kodas om till en textvariabel.



Vi avslutar med att fylla i menyrutorna du ser nedan till höger. Data kommer från kolumn C2 dvs Kön och ska till C7 som är första lediga kolumn.

”Original value” 1 kodas om till det nya värdet Man och 2:orna ska kodas till Kvinna.

Ett avslutande klick på OK och en ny textkolumn (lägg märke till att det står C7–T) har skapats i worksheeten.



	C1-T	C2	C3	C4	C5	C6	C7-T	C8	C9
	Namn	Kön	Födelseår	Motion	SMS	Ålder			
1	Lasse	1	1979	2	2	25	Man		
2	Bosse	1	1982	0	1	22	Man		
3	Lisa	2	1981	5	4	23	Kvinna		
4	Anna	2	1983	3	3	21	Kvinna		
5	Britta	2	1980	1	3	24	Kvinna		
6	Olle	1	1981	3	3	23	Man		
7	Kerstin	2	1985	2	2	19	Kvinna		
8									

En viktig skillnad gentemot kalkylprogram är att de nya värdena räknas ut och läggs som data eller text, och inte som en formel, i den nya kolumnen. Det innebär att om man i efterhand ändrar födelseåret så ändras inte åldern automatiskt.

3 Beskrivande statistik

Det första man bör göra med ett insamlat datamaterial är att beskriva det på olika sätt. Det datamaterial vi ska använda är från en enkät som 200 studenter besvarat. Worksheeten STUDENTER.MTW innehåller dessa variabler (kolumner). Filen finns på institutionens server på enheten M:

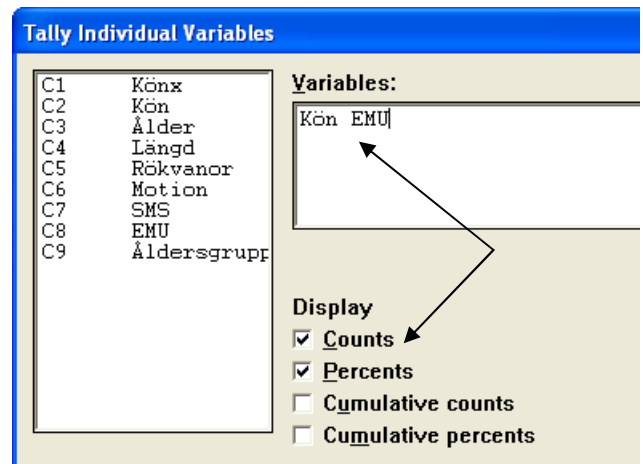
C1	Könx	1=Man 2=Kvinna
C2–T	Kön	Man / Kvinna
C3	Ålder	Ålder i år
C4	Längd	Kroppslängd i cm
C5	Rökvanor	1=Aldrig 2=Ibland 3=Ofta
C6	Motion	Antal motionspass per vecka
C7	SMS	”Hur ofta skickar du SMS?” 1=Aldrig 2=Någon gång/vecka 3=Flera gånger/vecka 4=Dagligen
C8	EMU	”Hur röstade du vid folkomröstningen om ett svenskt EMU-medlemskap?” 1=Nej 2=Ja *=Missing value
C9–T	Åldersgrupp	Yngre (–24 år) / Äldre (25 år +)

3.1 Frekvenstabeller

För att få enkla frekvenstabeller väljer man **Stat > Tables > Tally**. Markera och välj (Select) de variabler som ska tabelleras. Vi har även kryssat för Counts (frekvens, antal) och Percents. Ett klick på OK och tabellerna skrivs ut i Session-fönstret.

Vi ser av tabellen nedan att av de 200 som svarat är 90 kvinnor (45 %) och 110 män (55 %). Att kvinnor alltid kommer före män i tabeller beror inte på Minitabs artighet utan på att K kommer före M i alfabetet!

Vi ser också att hela 76,8 % röstade Ja till EMU – men också att 19 av de 200 inte röstade alls.



Tally for Discrete Variables: Kön; EMU

Kön	Count	Percent	EMU	Count	Percent
Kvinna	90	45,00	1	42	23,20
Man	110	55,00	2	139	76,80
N=	200		N=	181	
			*=	19	

3.2 Korstabeller

I korstabeller tabellerar man två kategorivariabler mot varandra. Vi vill veta hur många av kvinnorna respektive männen som röstade Ja eller Nej till EMU.

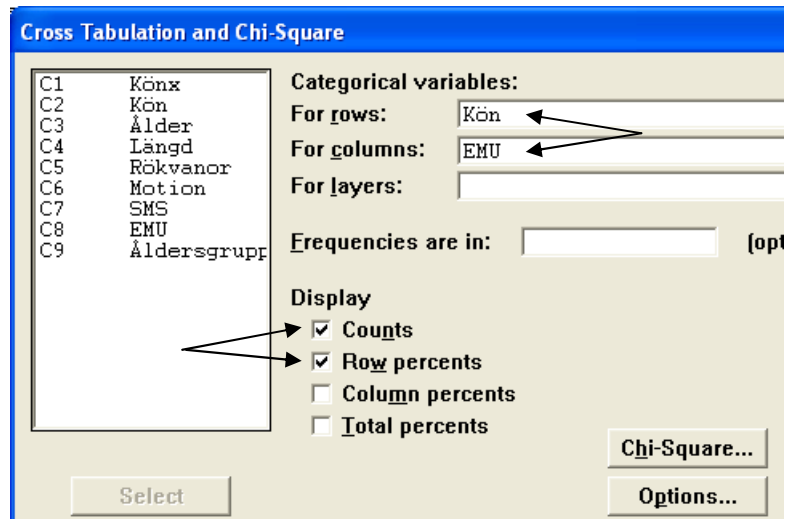
Stat > Tables > Cross Tabulation

leder oss till den här menyn.

Under "Categorical Variables:" ska vi nu bestämma vilken kategorivariabel som ska utgöra rader och vilken som ska utgöra kolumner. Här har vi valt Kön "For rows:" och EMU "For columns:"

När man beräknar procenttal i korsstabeller får man den intressantaste informationen om man gör det radvis eller kolumnvis. Här vill vi se hur många procent av männen respektive kvinnorna som röstade Ja och Nej. Därför väljer vi att kryssa för även "Row percents".

Eftersom kön är radvariabel får vi de önskade procenten på detta sätt. **OK** och här är resultatet.



Tabulated statistics: Kön; EMU

Rows: Kön Columns: EMU

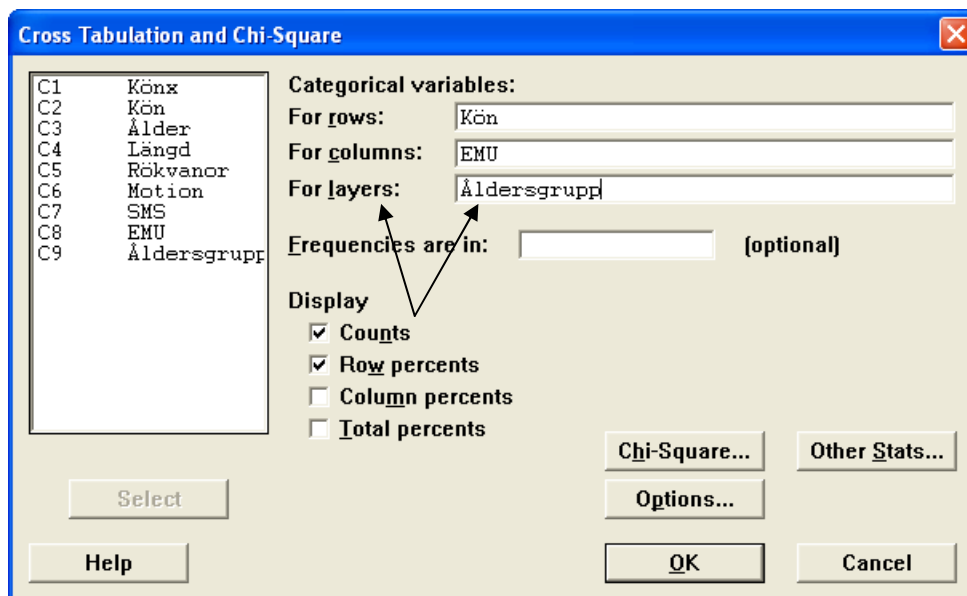
	1	2	Missing	All
Kvinna	26 32,10	55 67,90	9 *	81 100,00
Man	16 16,00	84 84,00	10 *	100 100,00
All	42 23,20	139 76,80	* *	181 100,00

Cell Contents: → Count
 % of Row

Vid "Cell Contents" står det vad varje "ruta" i korstabellen innehåller – "Count" betyder frekvenser och "% of Row" innebär just att procenten räknats ut varje rad (dvs kön) för sig. Av de 81 kvinnorna röstade 55 eller 67,9 % Ja till EMU. Bland männen var det hela 84 %.

Tycker man det är störande att kolumnen med "Missing" finns med ska man innan man klickar på **OK** klicka på knappen **Options** och där välja bort Missing values.

Det finns en utmärkt möjlighet att även ta hänsyn till en tredje kategorivariabel i Minitabs korstabeller. Man låter den tredje variabeln utgöra olika lager ("Layers").



På det här sättet får man en korstabell för varje värde på den tredje variabeln, dvs en korstabell för de yngre (upp till 24 år) och en för de äldre (minst 25 år).

Både ibland de äldre och yngre ser vi att män är mer positiva till EMU än vad kvinnorna är – 87 % mot 68 % bland de yngre och 80 mot 67 bland de äldre.

En annan tendens vi kan se här är att de yngre verkar vara något mer positiva till EMU än de äldre.

En liten varning kan vara befogad mot att använda alltför många variabler och värden i korstabellerna. De blir ganska snart svåra att läsa och tolka om det är för många siffror i dem.

Results for Åldersgrupp = Yngre

Rows: Kön	Columns: EMU			
	1	2	Missing	All
Kvinna	18 31,58	39 68,42	7 *	57 100,00
Man	8 13,33	52 86,67	8 *	60 100,00
All	26 22,22	91 77,78	* *	117 100,00

Cell Contents: Count
 % of Row

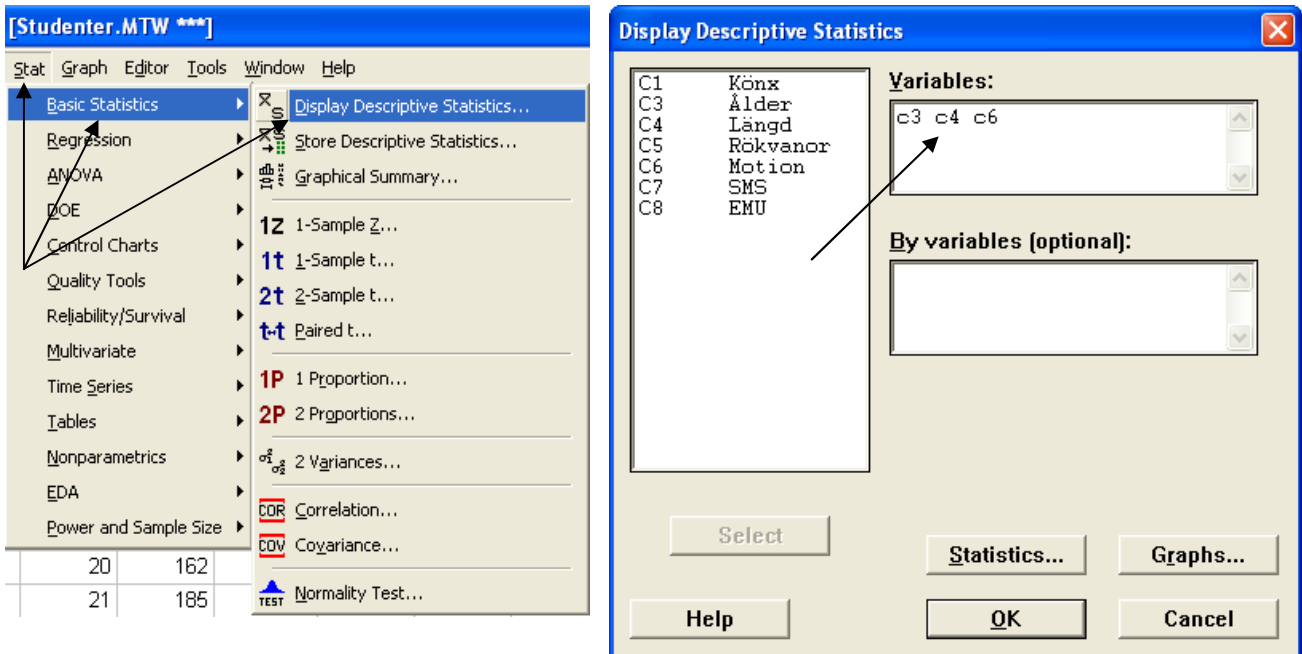
Results for Åldersgrupp = Äldre

Rows: Kön	Columns: EMU			
	1	2	Missing	All
Kvinna	8 33,33	16 66,67	2 *	24 100,00
Man	8 20,00	32 80,00	2 *	40 100,00
All	16 25,00	48 75,00	* *	64 100,00

Cell Contents: Count
 % of Row

3.3 Medelvärden å sånt

Stat > Basic Statistics > Display Descriptive Statistics är inledningen när man vill beräkna medelvärde, standardavvikelse, median, kvartiler och andra deskriptiva mått. Menyn till höger ska sedan fyllas i. Analysvariablerna skrivs eller klickas in vid pilen och ...



... det här resultatet visas när du klickat på **OK**. Den genomsnittlige studenten är 24,0 år gammal, 176,7 cm lång och motionerar 2,5 ggr per vecka.

Descriptive Statistics: Ålder; Längd; Motion

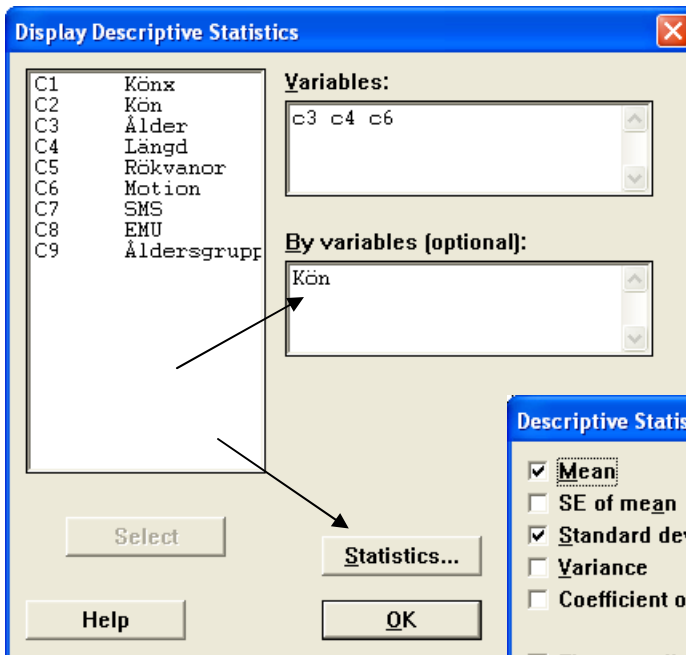
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Ålder	200	0	24,010	0,272	3,850	19,000	21,000	23,000	26,000
Längd	200	0	176,72	0,660	9,33	150,00	170,00	177,00	184,00
Motion	200	0	2,530	0,119	1,683	0,000000000	1,000	2,000	3,000

Variable	Maximum
Ålder	38,000
Längd	203,00
Motion	7,000

Att män i genomsnitt är längre än kvinnor vet vi men finns någon skillnad mellan könen i ålder eller motionsfrekvens? På nästa sida visar vi hur man kan få beräkningarna uppdelade på undergrupper.

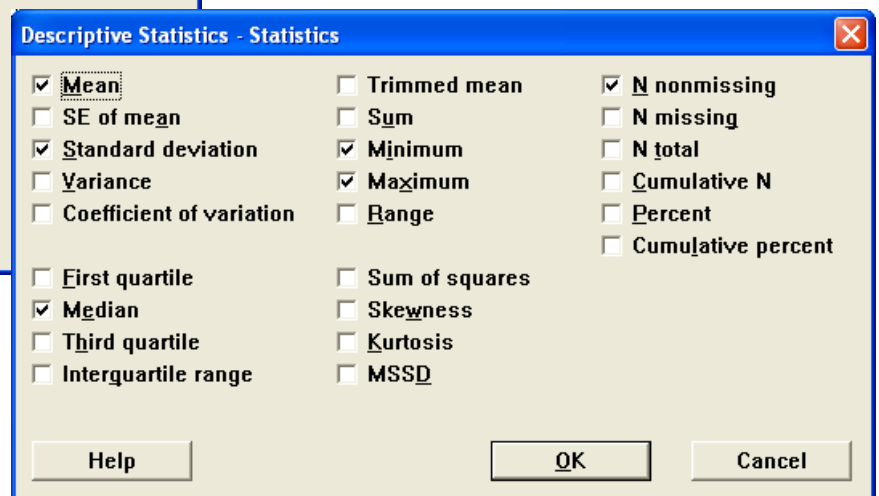
Vi visar också hur man bestämmer vilka värden som ska beräknas. Det är lite för mycket här ovanför. Vet du förresten vad alla olika värden innebär? Vad är t.ex. "SE Mean"? Man kan få en förklaring om man klickar på **Help**-knappen i menyn ovan – eller så frågar man sin kunnige lärare som gärna delar med sig av sin vishet!

3.4 Beskrivande statistik uppdelat på undergrupper



För att få beräkningarna uppdelade på undergrupper ska man klicka in gruppvariabeln under "By variables:".

Klickar vi sedan också på **Statistics**-knappen kan vi själva bestämma vilka mått som ska beräknas genom att kryssa för eller ta bort lämpliga kryss i menyn härunder.



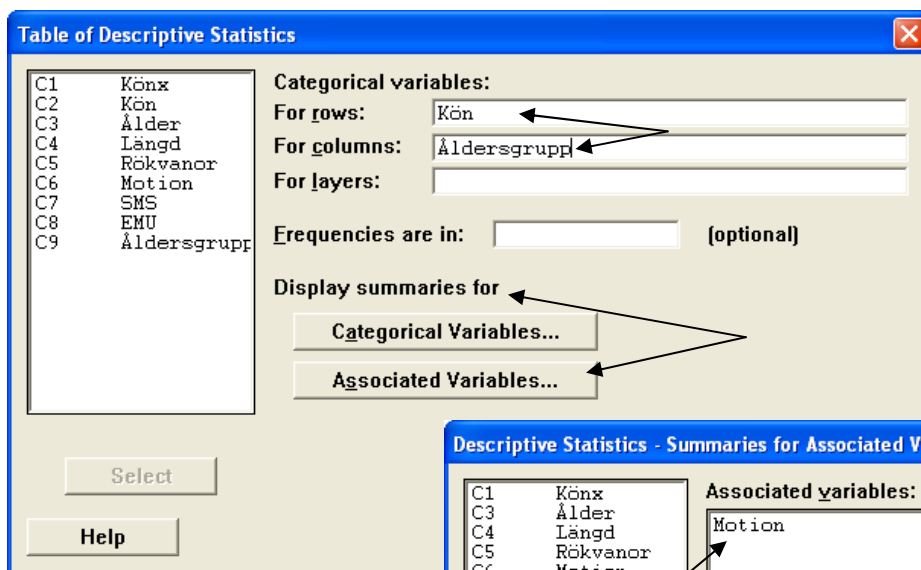
Det var väldigt många olika mått som kan beräknas. Klicka på **Help** så kommer förklaringarna i bildskärmen.

Resultatet av jämförelsen mellan könen ser du längst ner på sidan.

Kvinnorna är i snitt 169,1 cm och männen 182,9 cm. Standardavvikelsen är drygt 6 cm för båda grupperna. När det gäller motionsvanorna är skillnaden liten, 2,7 ggr för männen mot 2,4 ggr för kvinnorna. Är skillnaden statistiskt säkerställd? Hur man avgör det får ni lära er senare under kursen.

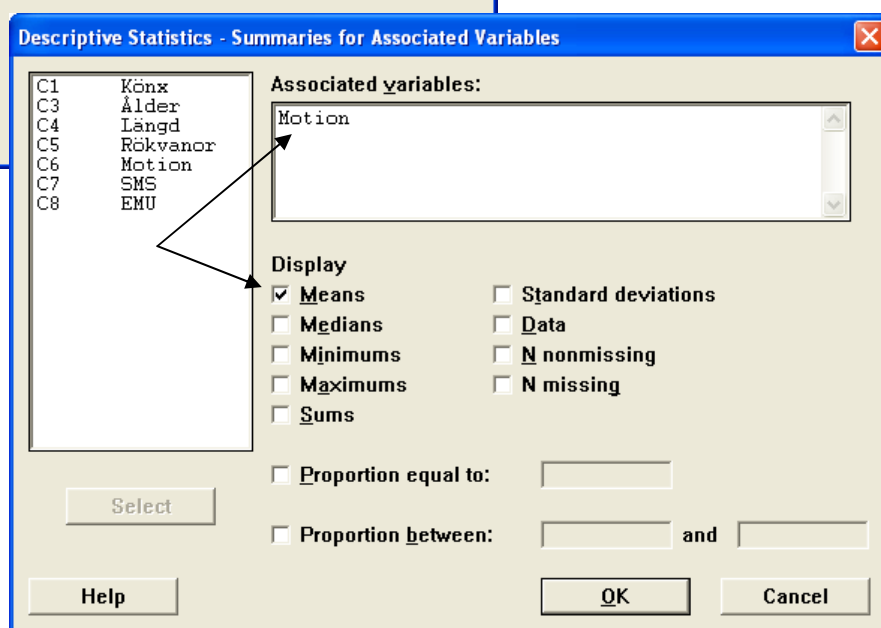
Descriptive Statistics: Ålder; Längd; Motion							
Variable	Kön	N	Mean	StDev	Minimum	Median	Maximum
Ålder	Kvinna	90	23,700	3,649	19,000	23,000	37,000
	Man	110	24,264	4,006	19,000	23,000	38,000
Längd	Kvinna	90	169,12	6,28	150,00	170,00	185,00
	Man	110	182,94	6,33	165,00	183,00	203,00
Motion	Kvinna	90	2,356	1,546	0,000000000	2,000	7,000
	Man	110	2,673	1,782	0,000000000	3,000	7,000

Det finns ytterligare ett sätt att få beskrivande statistik uppdelat på undergrupper och det är med **Stat > Tables > Descriptive Statistics**. Menyn påminner mycket om när man beställde korstabeller.



Här har vi valt att ha **Kön** som radvariabel och **Åldersgrupperna** som kolumner.

Därefter klickar vi på knappen **Associated Variables** och hamnar i nästa meny.



Som "Associated variable" väljer vi C6, Motion. Vi kryssar enbart för "Means" och får därmed medelvärden beräknade i den korstabell som kommer i Session-fönstret längst ner på sidan.

Tabulated statistics: Kön; Åldersgrupp

Rows: Kön Columns: Åldersgrupp

	Yngre	Äldre	All
Kvinna	2,500 64	2,000 26	2,356 90
Man	2,838 68	2,405 42	2,673 110
All	2,674 132	2,250 68	2,530 200

Cell Contents: Motion : Mean
Count

Ur den här tabellen utläser vi att yngre kvinnor motionerar 0,5 ggr mer per vecka än äldre, 2,5 mot 2,0.

För männen är värdena något högre men skillnaden mellan yngre och äldre är ungefär densamma, 2,8 mot 2,4.

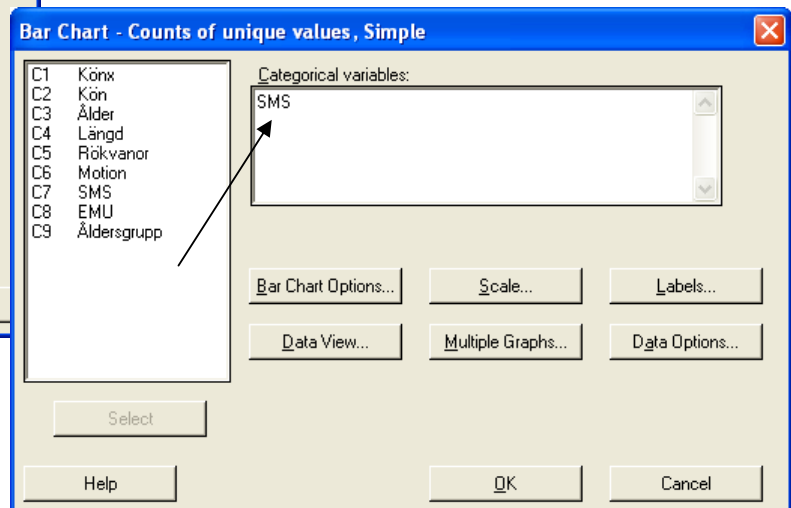
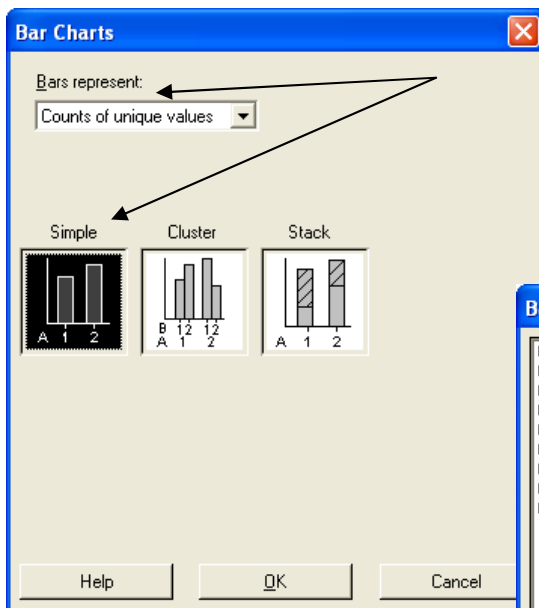
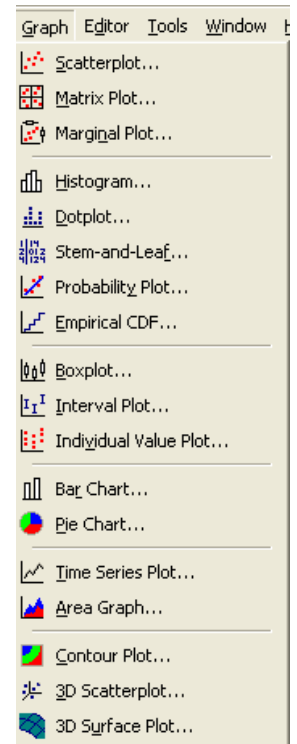
4 Några vanliga diagramtyper

Under Graph på menyraden döljer sig många olika diagramtyper. Vi ska enbart se på några av dem. Har man väl förstått principen för hur ett diagram redigeras fungerar det på samma sätt med de andra typerna.

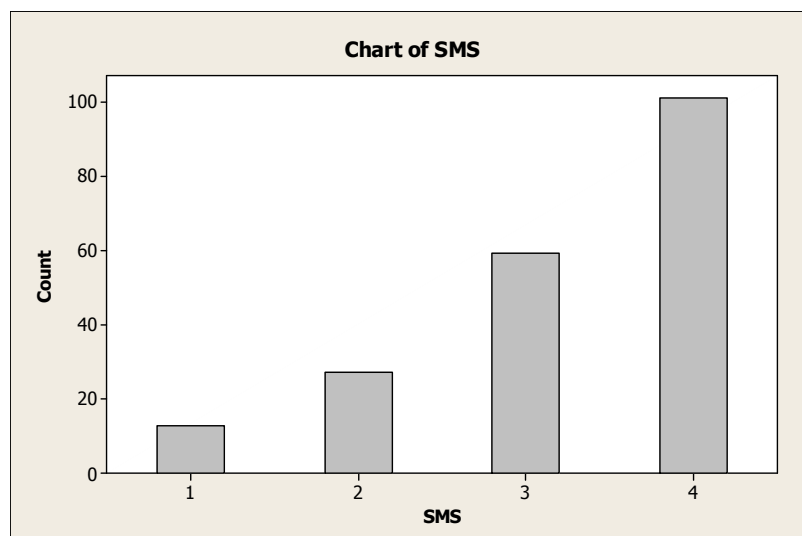
4.1 Stapeldiagram (Bar Chart)

Vi börjar med ett enkelt stapeldiagram. **Graph > Bar Chart** och en första meny dyker upp. Vi väljer att låta "Bars represent: Counts of unique values". Det betyder att Minitab i den aktuella kolumnen räknar efter hur många värden av varje slag det finns.

Simple ger oss ett enkelt stapeldiagram. I nästa meny väljer vi SMS-variabeln och efter **OK** får vi ett mycket enkelt stapeldiagram.



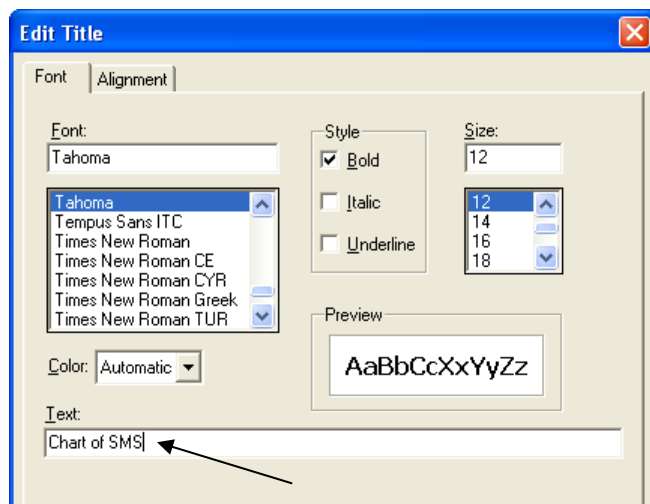
På nästa sida ska vi se hur diagrammet kan snyggas till.



Vi börjar med att ändra titeltexten över bilden. *Dubbelklicka* på texten "Chart of SMS" ovanför bilden så kommer den här menyn upp.

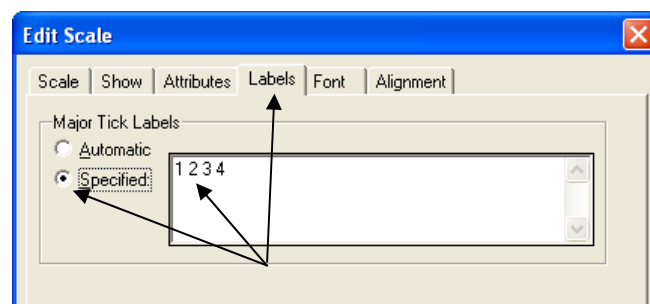
Ändra texten "Chart of SMS" till något lämpligare; "Studenters SMS-vanor" kanske. Vill man ändra font, färg och storlek på texten görs det också här.

På samma sätt ändrar man texten under och till vänster om bilden.



1, 2, 3 och 4 under staplarna är inte speciellt upplysande. Vi vill i klartext ha "Aldrig" etc till "Dagligen" i stället.

- Dubbelklicka på någon av siffrorna,
- välj fliken Labels,
- kryssa för "Specified",
- radera "1 2 3 4" och skriv i stället in "Aldrig Sällan Ofta Dagligen".

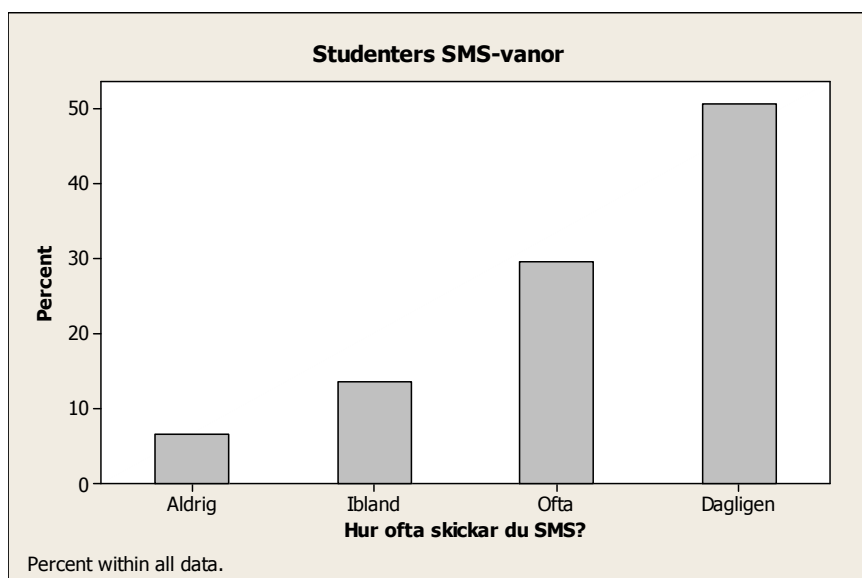
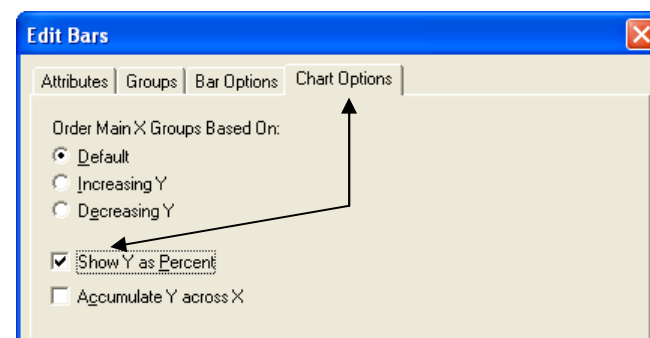


Avsluta med OK och vi är nästan färdiga.

Den sista ändringen den här gången blir att vi ändrar y-axeln från att ange antal individer till procent-tal. Dubbelklicka i någon av de fyra grå staplarna och ...

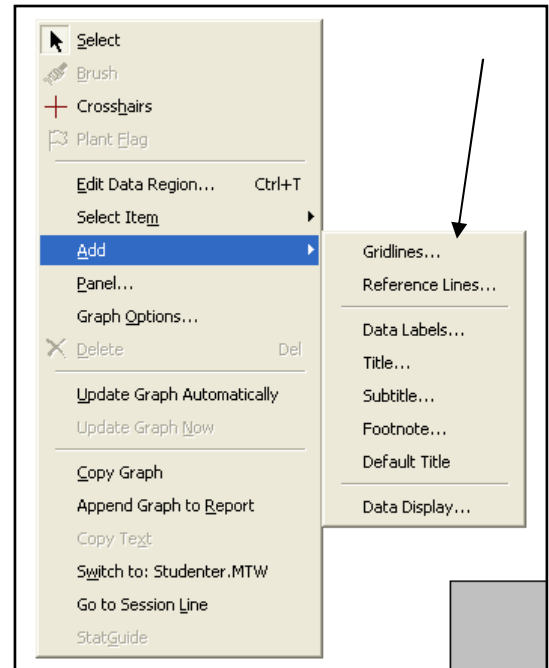
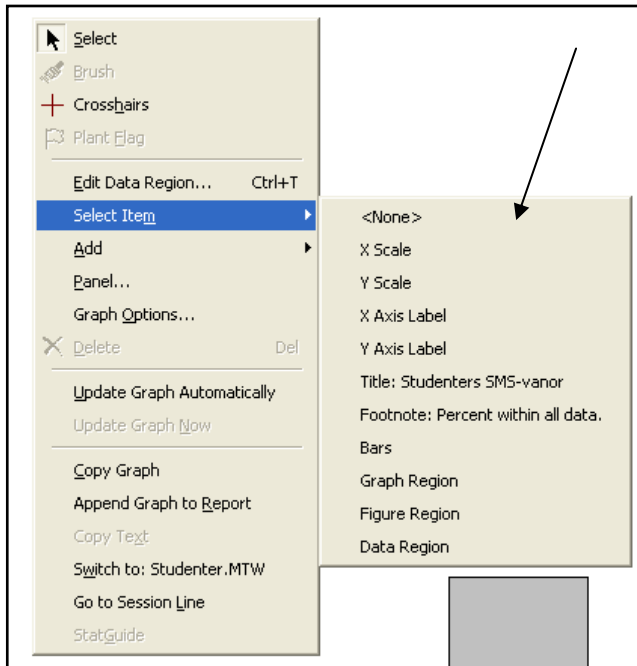
- välj fliken "Chart Options"
- och kryssa för "Show Y as Percent"

Nu är vi färdiga.



En annan möjlighet att redigera diagram är att redan innan man klickar OK i den första meny använda knapparna **Bar Options, Scale, Labels, Data View, Multiple Graphs, Data Options** för att bestämma bildens utseende.

Ytterligare ett alternativ är att klicka med *höger* musknapp mitt i bilden. Speciellt användbara alternativ är ”Select Item” och ”Add” – se nedan.

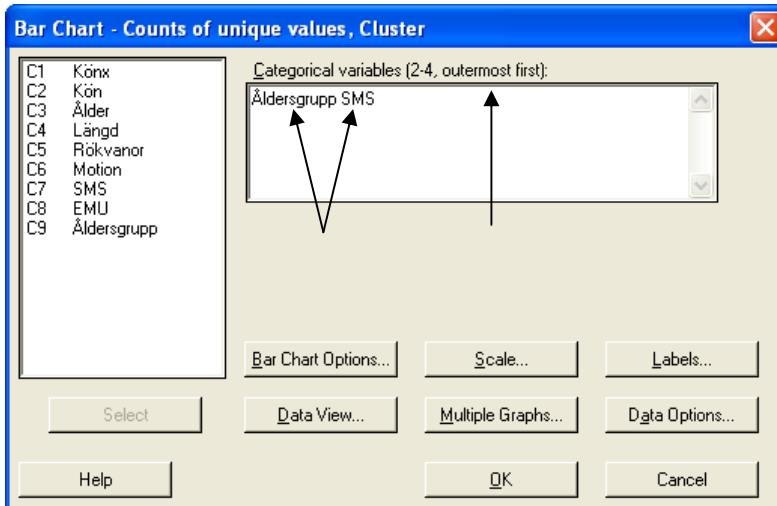
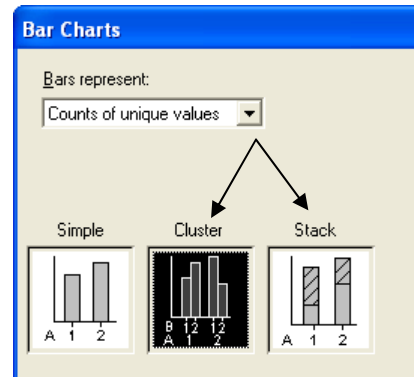


Att beskriva alla möjligheter i detta häfte är omöjligt utan man får helt enkelt testa sig fram. På de följande sidorna följer ytterligare några bildexempel – några redigerade och andra oredigerade.

4.2 Uppdelade eller grupperade stapeldiagram (Bar Chart)

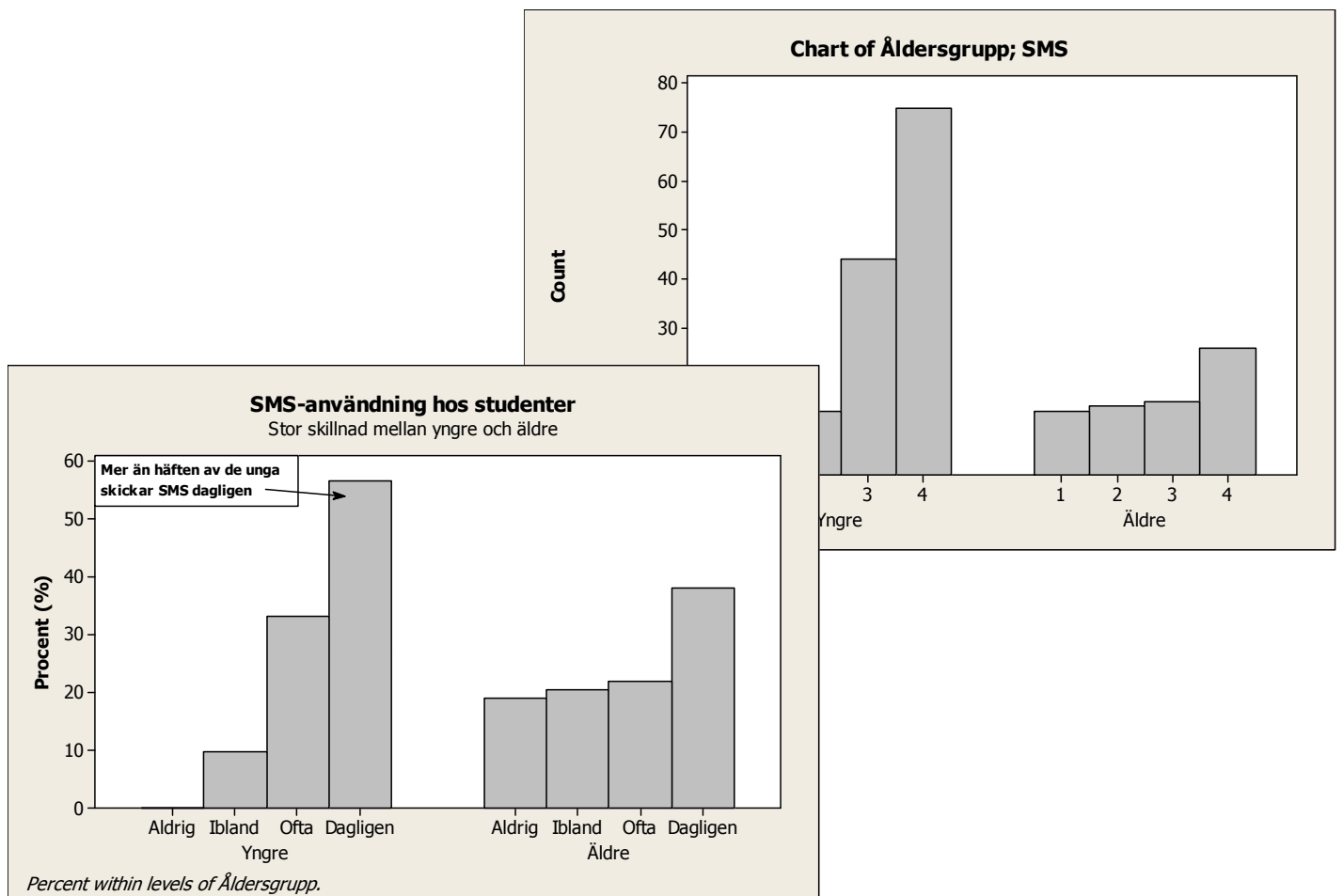
Finns den någon skillnad mellan äldre (över 25 år) och yngre studenter (under 25 år) när det gäller hur ofta de skickar SMS? Ett grupperat (clustered) eller uppdelat (stacked) stapeldiagram kan ge svaret?

Graph > Bar Chart > följt av den typ du vill ha. I exemplet har vi använt Cluster.



Åldersgrupp klickas in först och SMS därefter innebär att Åldersgrupp blir den "yttre" variabeln i figuren.

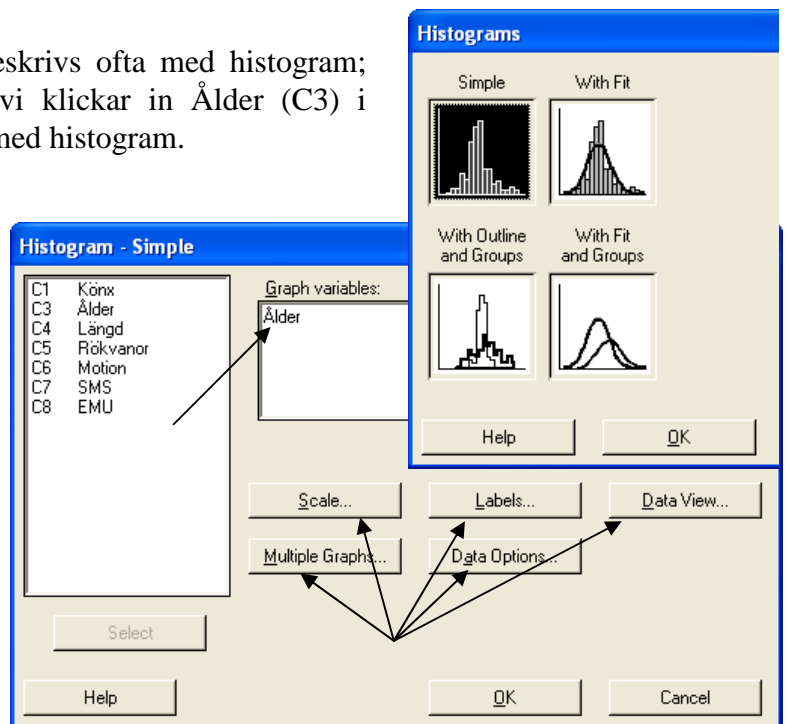
Nedan visas både den oredigerade och den redigerade bilden.



4.2 Histogram

Kontinuerliga variabler (ålder t.ex.) beskrivs ofta med histogram; **Graph > Histogram > Simple** och vi klickar in Ålder (C3) i rutan över variabler som ska beskrivas med histogram.

Med de olika knapparna markerade med pilar kan man redan från början snygga till diagrammet men vi väljer att göra det efter det att vi sett hur den första bilden blev.



Och den blev väl sådär

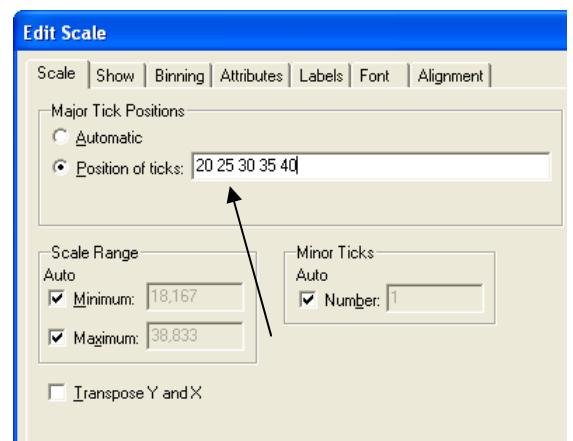


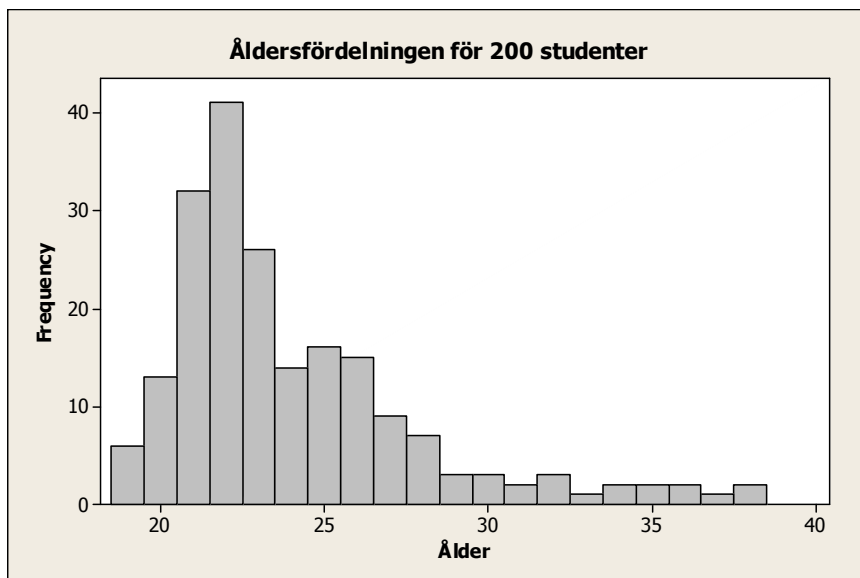
Vi ska ändra på

- rubriken (och hur man gör det vet vi redan)
- markeringarna längs x-axeln ska bytas från 21, 24 etc till 20, 25, 30 etc.

Börja med att dubbelklicka på rubriken och byt texten. Dubbelklicka därefter på något av värdena på x-axeln och ändra enligt här bredvid. Bocka för "Position of ticks" och byt ut 21 24 etc. mot 20 25 30 35 40. Ett kortare skrivsätt för detta är 20:40/5.

(forts till nästa sida)



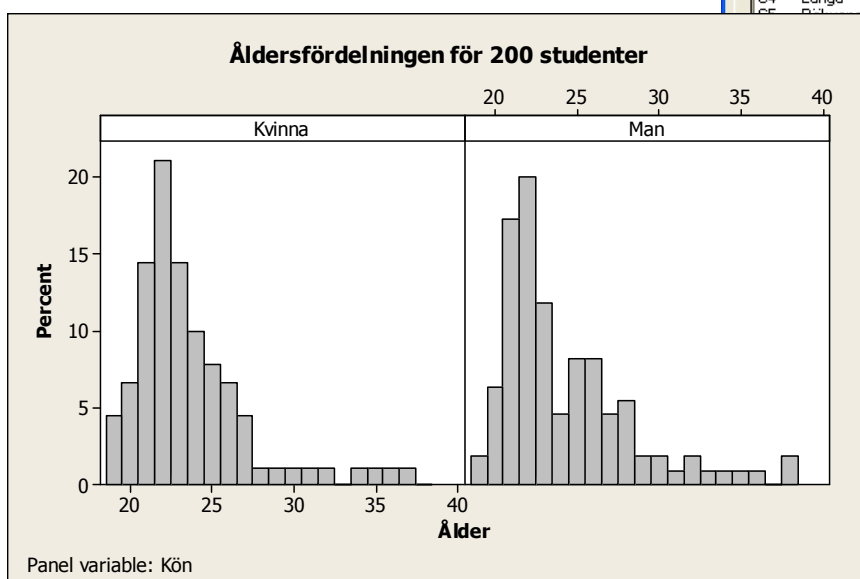
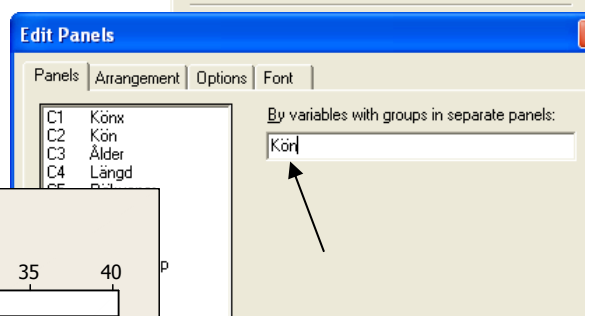
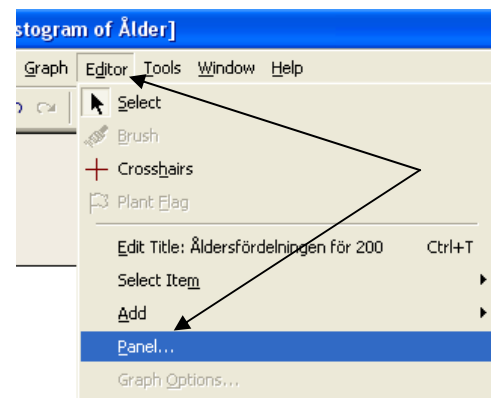


Som en liten extra finess kan man även i efterhand dela upp histogrammet på olika grupper för att se om åldersfördelningen är densamma.

Gör så här när du har bilden ovan i skärmen. Klicka på menyraden på **Editor > Panel**. Och i den följande menyn väljs **Kön** som variabel. Då får man histogrammet uppdelat på variabeln kön.

Eftersom det inte var lika många män som kvinnor i datamaterialet har vi i figuren längst ner också räknat om skalan på y-axeln till procent.

Det görs så här: Högerklicka på siffrorna vid y-axeln, välj "Edit Y-scale" och ändra till "Percent" under fliken "Type".

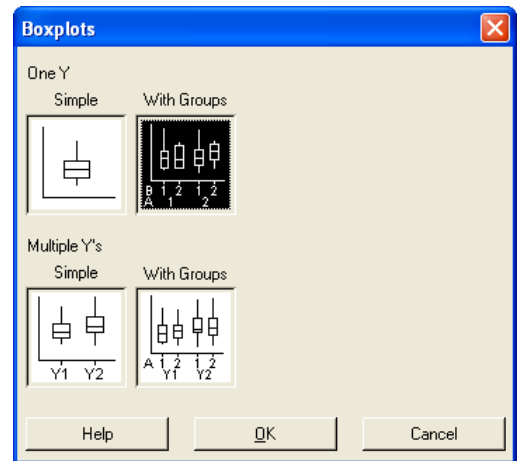


4.3 Boxplot

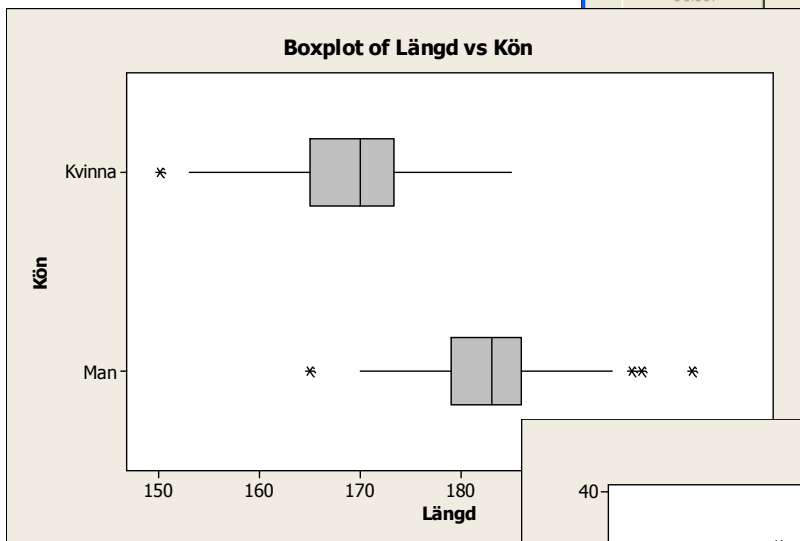
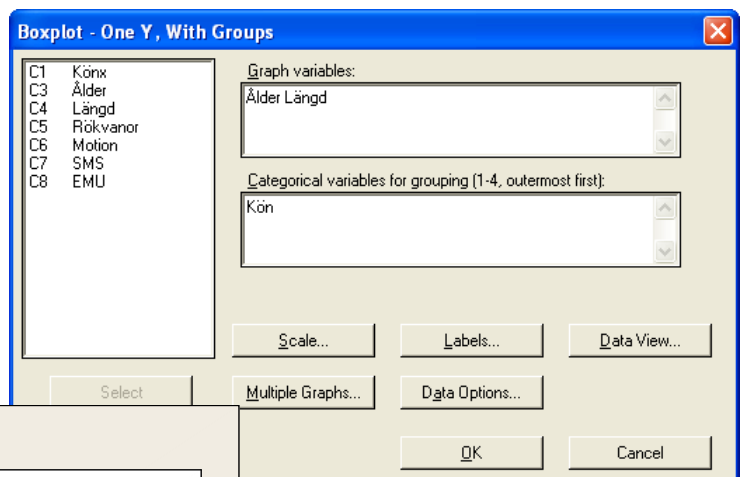
Boxplot är en utmärkt figur när man ska jämföra två eller flera gruppers värden på kontinuerliga variabler.

Graph > Boxplot och välj "With Groups", när det är gruppjämförelsen som är viktig.

I nästa meny väljer vi Ålder och Längd som analysvariabler och Kön som kategorivariabel. Man kan använda flera kategorivariabler samtidigt.

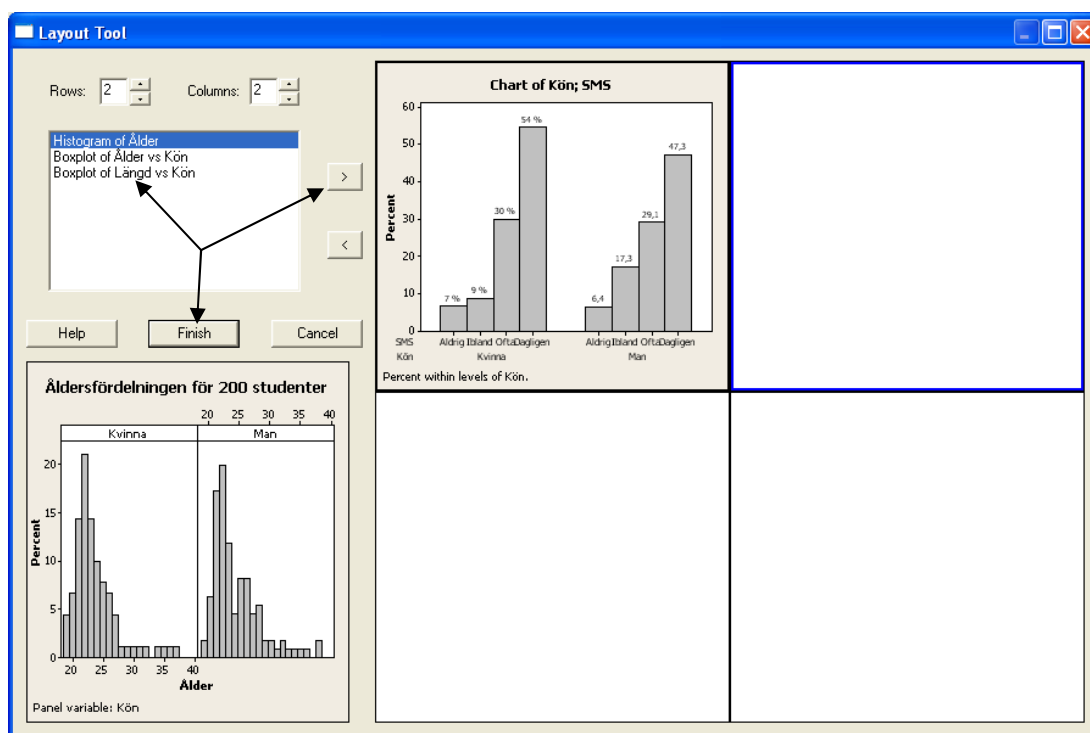


I figurerna nedan, som är lätt redigerade, ses en tydlig skillnad i kroppslängd men ingen åldersskillnad.



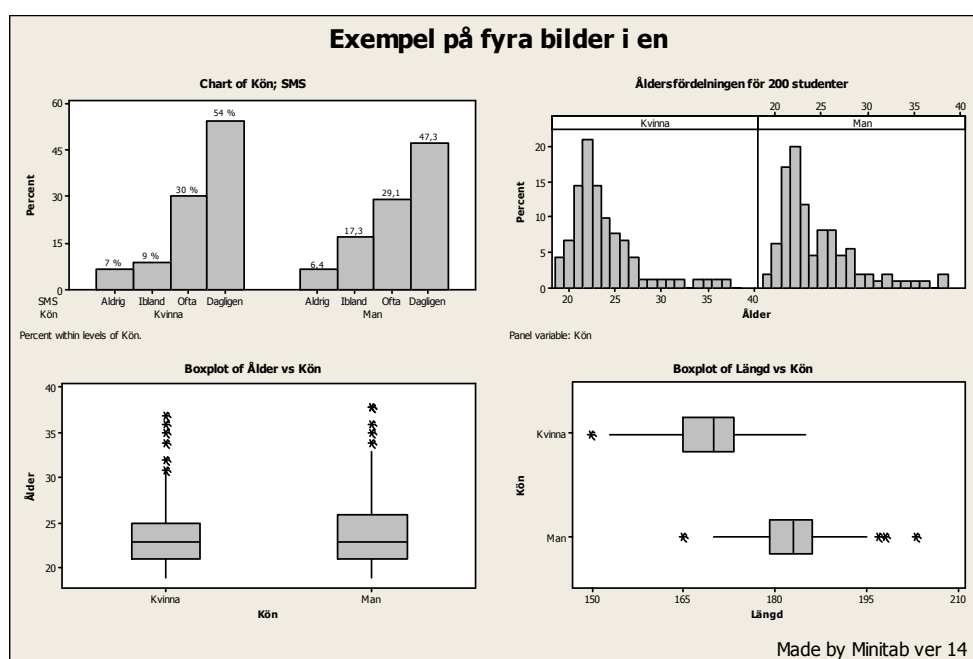
4.4 Layout-verktyget

Fyra av de diagram som ritats kan nu med hjälp av layout-verktyget plockas samman till en enda figur. Leta i Minitab rätt på diagrammet på sidan 17 och välj därefter **Editor > Layout Tool**.



Till höger ser man det utrymme där bilder kan läggas i 2 rader och 2 kolumner (eller ännu mer om man vill). Den aktuella bilden hamnar i övre vänstra hörnet. I listan över bilder till vänster markeras nu bilder en i taget och flyttas över till den högra delen med den lilla pilen (>). När alla fyra bilderna lagts över klickar man på Finish-knappen och en ny bild bestående av fyra små har skapats.

Editor > Add > Title resp **Footnote** och här är resultatet.



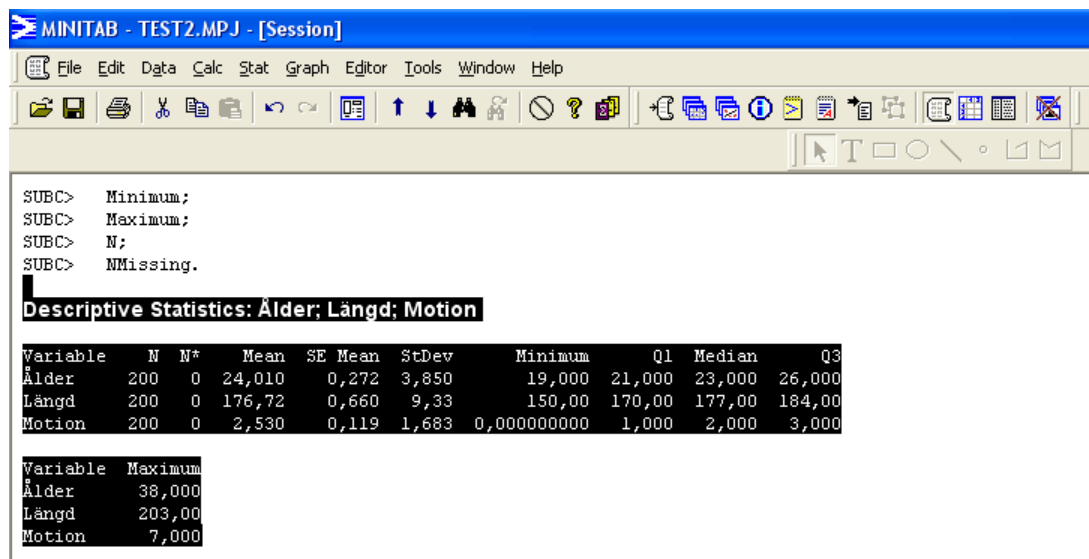
5. Hur för man över resultaten till Word?

Resultaten kopieras enkelt till Word. Tekniken är något olika beroende på om det är en grafisk bild eller om det är en del av Session-fönstret som ska flyttas till ordbehandlaren.

5.1 Kopiering av diagram från grafik-fönster

När den grafiska bilden är aktuellt fönster väljer man antingen **Edit > Copy Graph** eller **Ctrl+C**. När man sedan klistrar in bilden i Word ska man för att få högsta kvalitet helst välja **Redigera > Klistra in special > Bild**.

5.2 Kopiering från Session-fönstret



MINITAB - TEST2.MPJ - [Session]

File Edit Data Calc Stat Graph Editor Tools Window Help

SUBC> Minimum;
SUBC> Maximum;
SUBC> N;
SUBC> NMissing.

Descriptive Statistics: Ålder; Längd; Motion

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Ålder	200	0	24,010	0,272	3,850	19,000	21,000	23,000	26,000
Längd	200	0	176,72	0,660	9,33	150,00	170,00	177,00	184,00
Motion	200	0	2,530	0,119	1,683	0,000000000	1,000	2,000	3,000

Variable	Maximum
Ålder	38,000
Längd	203,00
Motion	7,000

Så här såg det ut i bildskärmen när resultaten på sidan 11 fixades till. Den del som skulle kopieras markerades med musen, **Ctrl+C** eller **Edit > Copy** och i Word kopierades med **Ctrl+V** eller **Redigera > Klistra in**. Kom sen ihåg att behålla teckensnittet Courier för att tabellen ska bli läsbar. Använder man ett proportionellt typsnitt som t.ex. Times New Roman nedan blir tabellen ”vinglig” och svåräst.

Descriptive Statistics: Ålder; Längd; Motion

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Ålder	200	0	24,010	0,272	3,850	19,000	21,000	23,000	26,000
Längd	200	0	176,72	0,660	9,33	150,00	170,00	177,00	184,00
Motion	200	0	2,530	0,119	1,683	0,000000000	1,000	2,000	3,000

Variable	Maximum
Ålder	38,000
Längd	203,00
Motion	7,000

OBS

Så här ska det alltså *inte* se ut. Tabellen på sidan 11 är bättre – där är tabellen i teckensnittet Courier.

6. Konfidensintervall och hypotesprövning

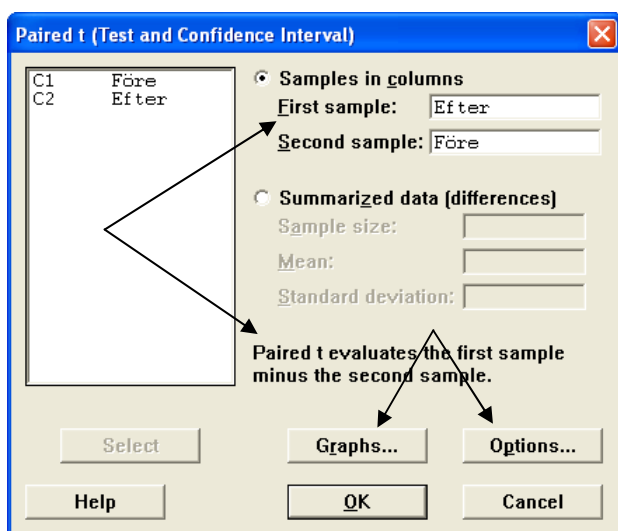
Under **Stat > Basic Statistics** hittar man statistisk analys av medelvärden och procenttal i ett eller två stickprov.

6.1 Medelvärden – parvisa observationer

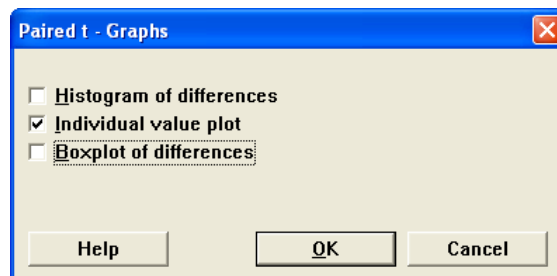
Uppgift 813 i Statistisk dataanalys handlar om analys av ett datamaterial där man mätt personernas kunskaper före och efter en kurs.. Data matas in i två kolumner (=variabler) och åtta rader, en rad per individ.

Vi vill testa om det finns någon genomsnittlig skillnad mellan ”Före” och ”Efter”-värdena. Eftersom vi har parvisa mätningar görs det med ett parvist t-test; **Stat > Basic Statistics > Paired t**.

	C1	C2
	Före	Efter
1	87	94
2	108	105
3	89	75
4	110	114
5	124	121
6	169	174
7	190	193
8	99	104
9		



När man klickat in sina analysvariabler vid First och Second sample kan man under knappen **Graphs** beställa några olika figurer och under **Options** bli bestämda om man vill ha ett ensidigt eller tväsidigt test. Tväsidigt (”not equal”) är standard. Man kan också ändra konfidensgraden för konfidensintervallet till något annat än det vanliga 95 %.



Så här blev resultatet i Session-fönstret:

Paired T-Test and CI: Efter; Före

Paired T for Efter - Före

	N	Mean	StDev	SE Mean
Efter	8	122,500	40,380	14,277
Före	8	122,000	37,834	13,376
Difference	8	0,500000	6,928203	2,449490

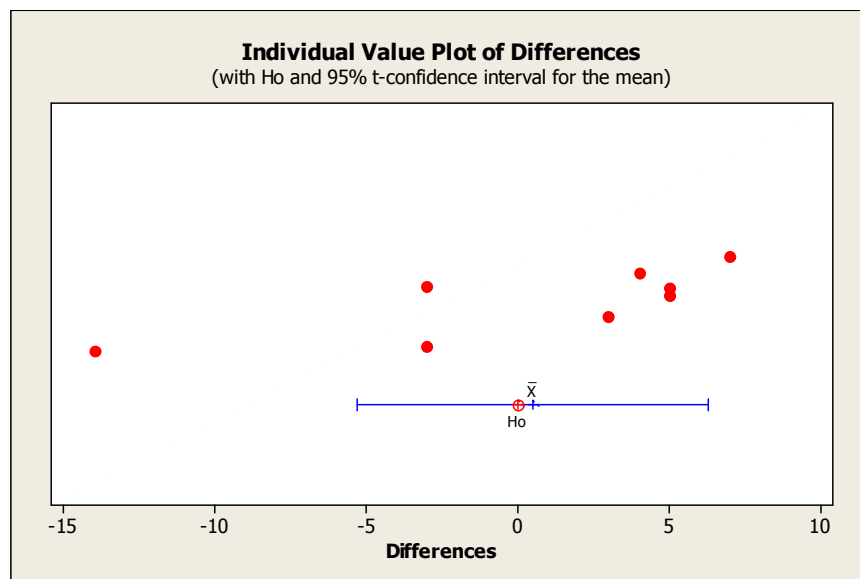
95% CI for mean difference: (-5,292123; 6,292123)

T-Test of mean difference = 0 (vs not = 0): T-Value = 0,20 P-Value = 0,844

P-värdet blev 84,4 % (0,844) vilket är långt över 5 %. Detta medför att vi accepterar nollhypotesen. Vi kan alltså *inte* påvisa att kursen förändrar kunskapsnivån.

Här är den ”Individual Value Plot” som vi också beställde. Nollhypotesens värde ($H_0 : \mu_{diff} = 0$) är markerat tillsammans med konfidensintervallet. Eftersom värdet noll ligger inom konfidensintervallet accepterar vi nollhypotesen – vilket vi redan konstaterat längst ner på föregående sida.

De åtta differenserna (Efter – Före) är också uträknade och inritade i figuren.

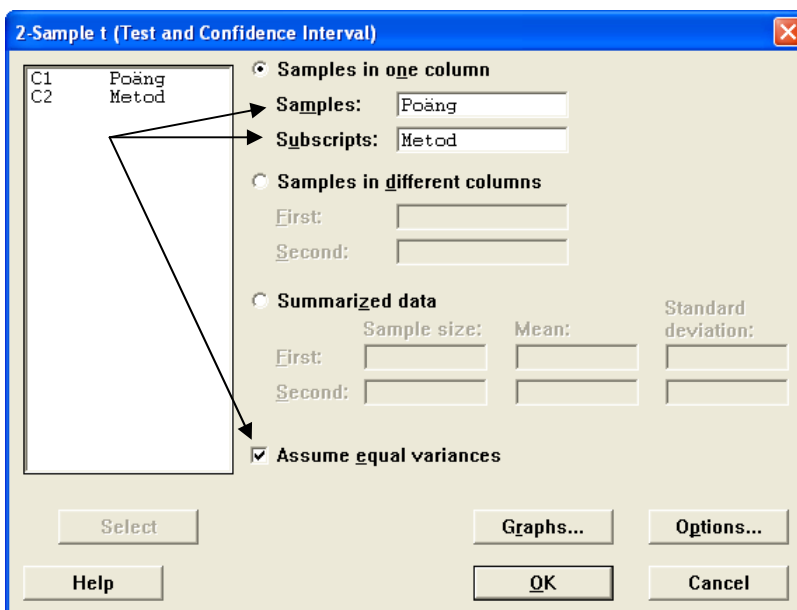


6.2 Medelvärden – två oberoende stickprov

I Statistisk dataanalys uppgift 809 analyseras ett datamaterial bestående av två oberoende stickprov. Vi har en analysvariabel, Poäng, och en gruppvariabel, Metod, som delar in personerna i de två grupper vi vill jämföra. Detta gör vi med

Stat > Basic Statistics > 2-Sample t ...

↓	C1	C2	C3
	Poäng	Metod	
1	29	1	
2	28	1	
3	18	2	
4	25	1	
5	25	2	
6	20	2	
7	25	1	
8	24	1	
9	24	1	
10	19	2	



I nästa meny ska vi berätta vilken variabel som är analysvariabel och vilken som är gruppindelingsvariabeln.

Vid ”Samples:” klickar man in analysvariabeln och gruppvariabeln vid ”Subscripts:” som egentligen betyder index.

Sen har vi ju också lärt oss att för att ett t-test ska vara korrekt i denna situation måste man anta att de båda grupperna har samma varians (SD sid 209). Detta gör vi genom att klicka i rutan ”Assume equal variances”.

Under **Graphs** ser vi till att fixa boxplottar eller Individual Value Plot och **Options** fungerar som tidigare.

Jämför resultatet nedan med bokens facit! Är skillnaden statistiskt säkerställd?

I utskriften nedan har Minitab gjort ett två-sidigt test. "(vs not =)" betyder att mothypotesen är att differensen inte är lika med noll.

Resultatet av t-testet blev $t = 1,80$ som med 28 frihetsgrader (DF = degress of freedom) ger ett tvåsidigt p -värde på $0,083 = 8,3 \%$.

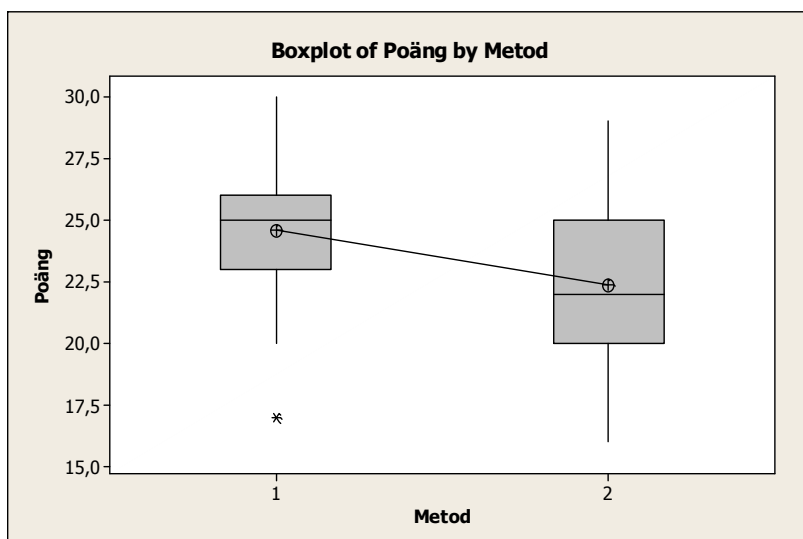
Men om man som i läroboken gör ett *ensidigt* test blir alltså p -värdet där bara hälften. $P = 0,083/2 = 0,0415 < 5 \%$ vilket medför att nollhypotesen förkastas. Det är statistiskt säkerställt att det nya programmet i genomsnitt ger ett högre resultat.

Two-Sample T-Test and CI: Poäng; Metod

Two-sample T for Poäng

Metod	N	Mean	StDev	SE Mean
1	15	24,60	3,20	0,83
2	15	22,40	3,50	0,90

Difference = mu (1) - mu (2)
 Estimate for difference: 2,20000
 95% CI for difference: (-0,30957, 4,70957)
 T-Test of difference = 0 (vs not =): T-Value = 1,80 P-Value = 0,083 DF = 28
 Both use Pooled StDev = 3,3552

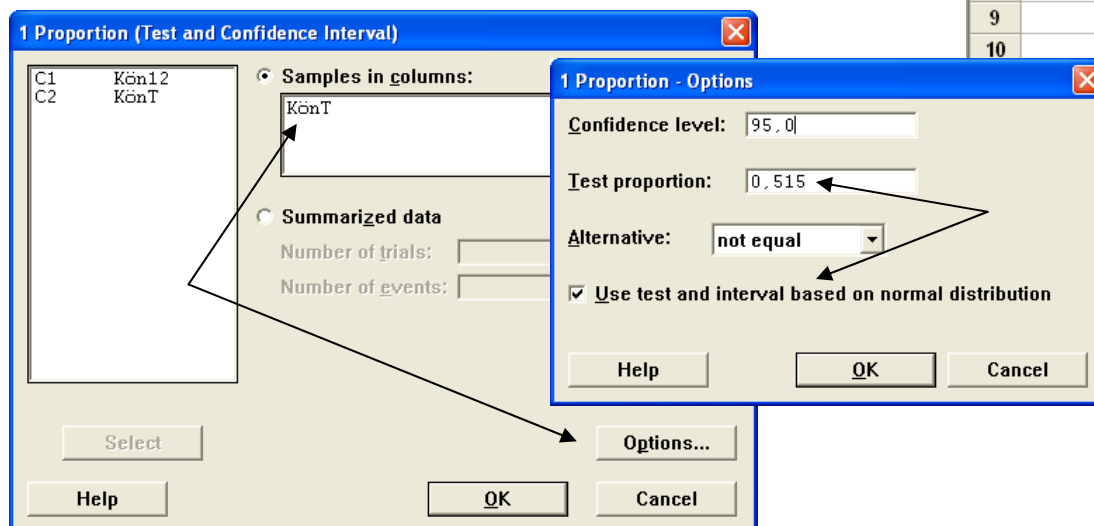


6.3 Analys av procenttal – ett stickprov

Får bestrålade män fler söner är frågan i uppgift 807 i Statistisk dataanalys. Datafilen ser du här bredvid; 345 rader med pojkar och flickor. Frågan är om andelen pojkfödslar avviker från det förväntade $106/206=0,515$. Det föds i genomsnitt 51,5 % pojkar och 48,5 % flickor i en vanlig population.

Stat > Basic Statistics > 1 Proportion ger menyn nedan.

↓	C1	C2-T	C3
	Kön12	KönT	
1	1	Pojke	
2	1	Pojke	
3	2	Flicka	
4	1	Pojke	
5	1	Pojke	
6	1	Pojke	
7	1	Pojke	
8	2	Flicka	
9	2	Flicka	
10	2	Flicka	



Under **Options** måste vi sedan ange att det är 0,515 som är värdet under nollhypotesen. Dessutom kryssar vi för att testet ska baseras på normalfördelningen på samma sätt som i läroboken.

Test and CI for One Proportion: KönT							
Test of $p = 0,515$ vs $p \text{ not} = 0,515$							
Event = Pojke							
Variable	X	N	Sample p	95% CI	Z-Value	P-Value	
KönT	202	345	0,585507	(0,533524; 0,637490)	2,62	0,009	

I Session-fönstret kommer resultatet. "Event = Pojke" står det. Det finns två värden i kolumnen och Minitab tar det högsta värdet om det är numeriska värden, dvs värdet 2 om vi hade valt C1. Är det textvärden tas det värde som kommer sist i alfabetisk ordning och P(ojke) kommer efter F(licka) i alfabetet.

Att 202 av 345 blir $0,5855 = 58,55\%$ får vi också reda på. Konfidensintervallet sträcker sig från 53,3 % till 63,7 % vilket lämnar "normalvärdet" 51,5 % utanför.

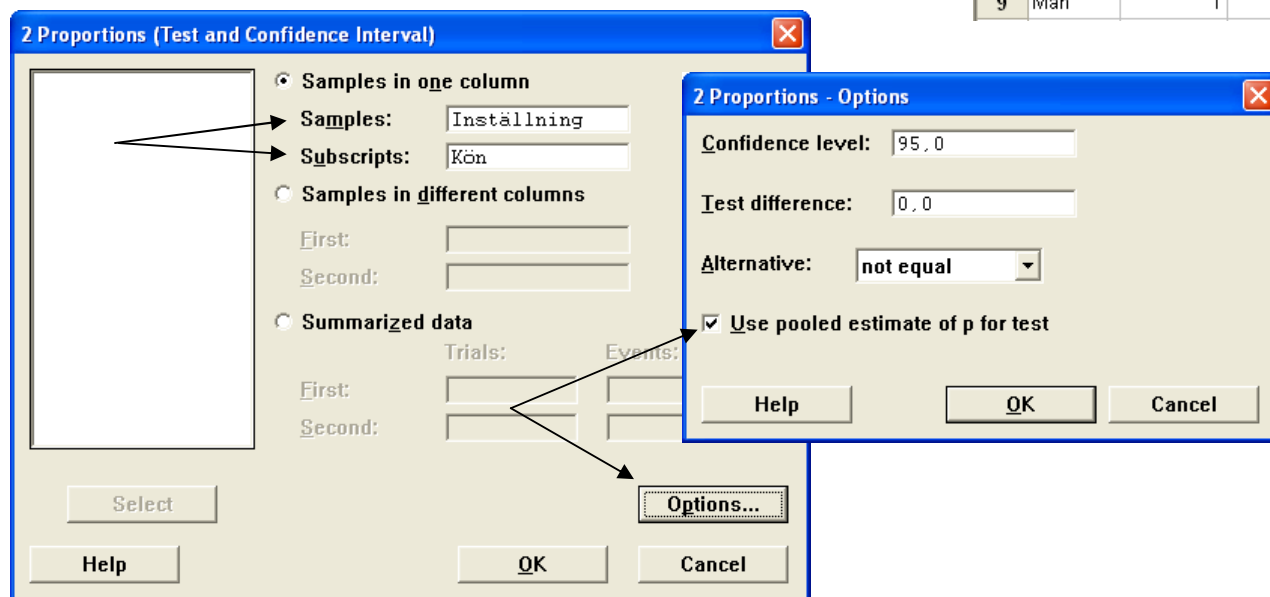
Slutligen utförs Z-testet och $Z = 2,62$ motsvarar ett *P-Value* på $0,009 = 0,9\%$ vilket innebär att vi har tvåstjärnig signifikans. Det är statistiskt säkerställt att "bestrålade män" får fler söner än normalpopulationen.

6.4 Analys av procenttal – två oberoende stickprov

I uppgift 808 i Statistisk dataanalys ställs frågan om det finns någon skillnad mellan männens och kvinnornas inställning till ett nytt arbetsmiljöprogram. Vid inmatningen av data har de positiva till förslaget kodats till värdet 1 och de andra till 2. Att jämföra två proportioner är i princip samma sak som att jämföra två medelvärden. I Minitab gör man det under **Stat > Basic Statistics > 2 Proportions ...**

Den meny som kommer upp är nästan identisk med den för medelvärdesjämförelser (sidan 25).

	C1-T	C2	C3
	Kön	Inställning	
1	Man	1	
2	Man	2	
3	Kvinna	1	
4	Man	2	
5	Kvinna	2	
6	Man	1	
7	Man	1	
8	Kvinna	2	
9	Man	1	



”Inställning” är analysvariabel och ”Kön” grupperna som jämförs. Under **Options** kryssar man för att testet ska göras med den sammanvägda proportionen (”Use pooled estimate of p for test”).

Test and CI for Two Proportions: Inställning; Kön			
Event = 2			
Kön	X	N	Sample p
Kvinna	46	100	0,460000
Man	37	100	0,370000
Difference = p (Kvinna) - p (Man)			
Estimate for difference: 0,09			
95% CI for difference: (-0,0460023; 0,2260023)			
Test for difference = 0 (vs not = 0): Z = 1,29 P-Value = 0,196			

”Event = 2” betyder att vi testar om andelen icke-positiva är lika stor bland kvinnor som bland män. Skillnaden mellan 0,46 och 0,37 är 0,09. Hade man jämfört andelen positiva hade det blivit 0,64 mot 0,73 vilket också är 0,09.

Testet ger värdet $Z = 1,29$ och $P\text{-value} = 0,196$. Ett p -värde på 19,6 % är så pass stort att nollhypotesen accepteras. Det finns alltså ingen signifikant skillnad mellan kvinnor och män – i detta avseende i alla fall.