

Datalaboration 1 - 732G01/732G40

HT2018

Introduktion

Denna datorövning behandlar visualisering av olika typer av variabler.

Övningen ska göras i Minitab. Minitab finns på datorerna i PC1-5 som vissa är bokade för dessa pass men Minitab finns också att ladda ner via Studentportalen både till Windows och Mac-datorer. Följ instruktionerna ni hittar där. Notera att Mac-versionen av Minitab ser något annorlunda ut och instruktionerna nedan är anpassade för Windows.

Datamaterial

Börja med att ladda ner det datamaterial som finns på kursens LISAM-sida. Datamaterialet innehåller tre flikar med information. Den första fliken behandlar data om längd och vikt hos 100 män och 100 kvinnor. Männerna är kodade med 2 och kvinnorna med 1. Den andra fliken behandlar information om civilstånd och bilmärken på Påhittade gatan. Den tredje fliken behandlar information om tentaresultat och ålder hos ett antal påhittade studenter.

För att importera datamaterialet till Minitab, starta programmet och gå till *File -> Open*. Sök upp den filen ni laddat ner och klicka på *Öppna*. Om allt gått rätt till har nu tre arbetsblad öppnats i Minitab med data från de tre flikarna.

Uppgift 1

Markera bladet **Längd** och **Vikt** och fortsätt med uppgifterna. I denna uppgift ska vi titta hur Minitab kan analysera och visualisera kontinuerliga kvantitativa variabler.

1. Börja med att beräkna lite deskriptiv statistik över datamaterialet. Gå till *Stat -> Basic Statistics -> Display Descriptive Statistics* och välj **Längd** och **Vikt** som variabler samt **Kön** som by.

Titta i utskriften och identifiera \bar{x} , s och andra relevanta mått på variablerna.

2. Nästa steg är att undersöka fördelningen av **Längd** och **Vikt**. Gå till *Graph -> Histogram -> With Groups* och välj variablerna som ska visualiseras samt **Kön** som kategorisk variabel. Innan vi skapar diagrammet klicka på *Multiple Graphs* och markera *In separate panels of the same graphs* för att de två variablerna ska bilda två separata diagram. I menyn *Scale -> Y-Scale Type* ska *Percent* markeras för att presentera **RELATIVA** frekvenser som är ett måste vid all form av visualisering. Klicka OK två gånger.

Vad kan vi säga om fördelningen av de olika variablerna? Skillnader mellan könen?

Uppgift 2

Markera nu bladet som heter **Bilmärke** och fortsätt med uppgifterna. Detta blad innehåller två stycken kvalitativa variabler.

1. Skapa en frekvenstabell över variablerna **Bil** och **Civilstånd** genom att gå till *Stat -> Tables -> Tally Individual Variables*. Markera därefter båda variablerna, och säkerställ att *Counts* är markerad under *Display*.

Vilket bilmärke är den vanligaste? Hur många har civilstånd "par"?

2. Att presentera tabeller är inte alltid det enklaste sättet för läsaren att få en bild av data. Det är oftast enklare (och snyggare) att skapa ett diagram.

I Minitab gör vi ett lämpligt stapeldiagram genom att gå till *Graph -> Bar Chart* och markera *Counts of unique values* och *Cluster*. Under *Chart Options* ska *Show Y as Percent* bockas för för att presentera relativa frekvenser. Markera **Bil** och **Civilstånd** och tryck OK.

Vilket är det vanligaste bilmärket för singlar respektive par?.

Uppgift 3

Markera nu det sista bladet som heter **Elevålder**. Detta blad innehåller två diskreta kvantitativa variabler.

1. Börja med att visualisera fördelningen av dessa två variabler. Här kan vi antingen välja ett histogram eller stolpdiagram på **Ålder** eftersom man kan tänka sig att värdena kan anta decimalvärden. För **Resultat** bör ett stolpdiagram (bar chart) väljas. Lägg märke till att stolpdiagrammen i Minitab inte följer en skala på x-axeln vilket gör det svårt att se avvikande observationer. Glöm inte att y-axeln ska visa relativa frekvenser!

Använd de skapade diagrammen och besvara:

- (a) Hur stor andel fick full pott, alltså 20 poäng, på tentan?
 - (b) Vilka åldrar är mest frekventa?
2. För att enklare visualisera avvikande observationer kan med fördel ett prickdiagram (Dotplot) skapas. Gå till *Graph -> Dotplot -> One Y Simple* och markera båda variablerna. Vad kan vi dra för slutsatser om avvikande observationer för de båda variablerna?
 3. Ibland kan man vilja dela upp en kvantitativ variabel, t.ex. ålder, i olika klassindelningar. Detta kallas med ett annat ord **diskretisering**.

Detta gör vi i Minitab genom att gå till *Data -> Recode -> To Text...*, väljer kolumner **Ålder** i *Recode values in the following columns* och under *Method*: väljer *Recode range of values*.

I kolumnen *Lower endpoint* skriver vi in 0, i *Upper endpoint* skriver vi in 20 (värdet strax över det som skall inkluderas i intervallet) och i *Recoded value* skriver vi in -19. Detta innebär att värdena 0 till 19 från kolumnen **Ålder** kommer heta "-19" i nya kolumnen. Fortsätt med att skapa 5-årsintervall fram till intervallet "35-39". Det sista intervallet ska då vara "40-". Namnge den nya variabeln till **Åldersklass**. Visualisera denna variabel på ett lämpligt sätt.