

Innføring i Excel

Et lite selv-instruksjons kurs (“tutorial”)

Den beste og raskeste måten å lære seg et nytt program på er på forhånd å ha en oppgave man ønsker å bruke programmet til å løse. Man lærer seg først det aller enkleste for å komme inn i programmet, ofte ved hjelp en “tutorial” som ofte er en del av hjelpe-funksjonen i mange programmer. Deretter lærer man det som skal til for å løse oppgaven ved hjelp av hjelpe-funksjonen i programmet, manualer hvis de fins tilgjengelige, fra Internet, kamerater eller lignende. Man må regne med litt fumbling og famling i begynnelsen, men man lærer fort på denne måten. De som har vært borti Excel før kan hoppe rett på oppgave 2, som er en slik oppgave, men der man blir leiet i hånden det meste av veien. De som aldri har brukt Excel kan begynne på oppgave 1.

Oppgave 1

Excel har en vel-utviklet hjelpe-funksjon, men ikke noen passende “tutorial” så vidt jeg kan se. Det fins imidlertid en masse fri “tutorials” på Internet. En del, både for nybegynnere og viderekommende, er samlet på

<http://math.about.com/od/excel/>

Slå opp på denne siden. Der finnes bl.a. en kort en på 4 leksjoner med navn ”Beginning Excel” og en litt mer utførlig, ”22 Fundamental Excel Lessons”. Bruk den dere liker best. Mens dere jobber med lære-programmet, prøv å ha læreprogrammet i et vindu og en åpnet Excel i et annet vindu på skjermen. Reproduer så det dere leser i læreprogrammet ved hjelp av Excel for å se at det virker!

Men bruk ikke for lang tid på dette! Begynn så fort dere føler at dere kan, på oppgave 2. Dere kan alltid vende tilbake til læreprogrammet når det trengs.

Oppgave 2

(Merknad: Når du åpner Excel for denne oppgaven kan det være lurt først å formatere hele arket som "Number", for å unngå rare ting som at tall man skriver inn endrer seg til dato, eller at visse formler ikke virker som de skal. Marker først hele arket med Ctrl+A. Gå så til menyen Format->Cells og marker "Number" på listen over formater som kommer fram.)

En skogbruker ønsker opplysning om mengden av tømmer i en gitt skog. Han deler først skogen opp i små jordstykker (plots) av omtrent samme størrelse og trekker ut tilfeldig 49 av dem som undersøkes grundig. For hvert jordstykk måles mengden av tømmer målt i favner (1 favn er ca. 3.2 m^3). Tallene er gitt under (*antakelig gjelder komma istedenfor punktum som desimal-skilte på pc-ene på pc-stua*)¹:

| | | | | | | |
|------|------|------|------|------|------|------|
| 39,3 | 14,8 | 6,3 | 0,9 | 6,5 | 3,5 | 8,0 |
| 10 | 1,3 | 7,1 | 6 | 17,1 | 16,8 | 0,7 |
| 7,9 | 2,7 | 26,2 | 24,3 | 17,7 | 3,2 | 7,4 |
| 6,6 | 5,2 | 8,2 | 5,9 | 3,5 | 8,4 | 44,8 |
| 8,5 | 13,4 | 19,4 | 19 | 14,1 | 1,9 | 12 |
| 19,7 | 10,3 | 3,4 | 16,7 | 4,3 | 1 | 7,6 |
| 28,3 | 26,2 | 31,7 | 8,7 | 18,9 | 3,4 | 10 |

- A. Legg disse 49 tallene inn i Excel i en enkelt kolonne, for eksempel i cellene B2 : B50, der kolon (:) betyr at alle cellene imellom skal være med. Skriv et navn på variabelen i B1 (for eksempel *X* eller *Volum* eller lignende). Du kan naturligvis skrive inn tallene i hver celle for hånd, men det letteste er faktisk å kopiere (med copy) tallene ovenfor og så lime inn i en kolonne - trykk for eksempel på celle B2 og deretter på *paste* under Edit-menyen, eller simpelthen Ctrl-V.

Symmetriske fordelinger er ofte lettere å håndtere enn skjeve. Det kan derfor være en fordel om et data-sett har en (tilnærmet) symmetrisk fordeling. Et symptom på at vi har en skjev fordeling er bl.a. når det er betydelig forskjell mellom gjennomsnittet og medianen. Du skal nå beregne gjennomsnittet og medianen for disse tallene på flere måter:

- (i) *Bruk av "Descriptive Statistics" fra modulen "Data Analysis"*. Hent inn dialogen "Descriptive Statistics" (fra menyen: Tools -> Data Analysis -> Descriptive Statistics).

Hvis Data Analysis ikke står på Tools-menyen, gå inn på Tools -> Add-Ins og klikk på den lille ruten foran "Analysis ToolPak".

I dialogen som kommer fram, angi først området der dataene ligger, i det lille vinduet ved *Input Range*. Du kan skrive inn området (for eksempel B2:B50), men det enkleste er å markere området med musen. Klikk først i vinduet slik at indikatoren | blinker der. Klikk så på første celle med data (B2). Gå deretter til siste celle (B50) med data og klikk på denne mens shift-knappen holdes nede. Da skal hele data-listen være markert, og i *Input Range*-vinduet står nå

¹ Du kan selv velge om du vil ha komma eller punktum som desimalskilte: Gå inn i menyen Tools -> Options -> International, og spesifiser komma eller punktum i vinduet for "Decimal separator".

\$B\$2:\$B\$50. Dollar-tegnene betyr at referansene er faste – dvs. ikke påvirket ved kopiering.

Gå så til *Output options* og bestem om du vil ha output i samme ark som data (*Output Range*) eller i et annet ark i samme bok (*New Worksheet*) eller i en ny bok.

Når du åpner en ny Excel-sesjon får du fram en såkalt bok bestående av tre tomme ark indikert nederst på skjermbildet. Du kan legge til nye ark i en bok etter behov ved hjelp av *Insert* på menyen.

Velg for eksempel *Output Range* og plasser indikatoren i det tilhørende vinduet. Klikk så på en celle i arket ditt der du vil at output skal starte. Referansen til denne cellen skal nå stå i *Output Range*-vinduet. Til slutt, merk av i den lille boksen for *Summary Statistics*. Ved deretter å trykke OK, kommer det fram en tabell over beskrivende størrelser for data, bl.a. gjennomsnitt (mean) og median.

- (ii) Beregn nå medianen slik det er gjort i Løvås: Sorter data først i stigende rekkefølge og plukk ut det midterste tallet. Lag først en kopi av data i kolonne D, for eksempel. Disse kan sorteres ved først å markere dem og deretter trykke på knappen på verktøylinja som er markert ved en A og Z over hverandre og en nedover-pil. Eventuelt fra menyen: *Data -> Sort*.

En tredje måte er å bruke Excel-funksjonen, MEDIAN. Velg en celle der du vil plassere medianen og klikk på denne. Gå så til menyen: *Insert -> function*. Velg *Statistical* i vinduet for category og klikk på MEDIAN i vinduet under. Til slutt klikk OK. Da kommer det fram en dialog. Hvis denne dekker noen dataene dine, kan du flytte den ved å peke på den og dra den vekk mens du holder venstre museknapp nede. Det eneste du trenger å gjøre nå er å indikere i vinduet for *Number 1* området for dataene, enten ved å skrive det inn direkte, eller markere området med pilen som beskrevet ovenfor. Trykker du på OK skal du i den cellen du startet fra, få samme median-verdi som du har funnet før.

Merk at når du klikker på den cellen der du har lagt medianen, framkommer selve formelen som har beregnet denne verdien, i et vindu øverst over arket. Denne formelen hadde du kunnet skrevet direkte inn i det vinduet uten å gå veien om dialogen først.

NB! Merk også at enhver formel i Excel starter med et likhetstegn, =.

En tredje ting å merke seg er at hvis data endrer seg, vil verdien for medianen også endre seg. Prøv for eksempel å endre den nest siste observasjonen, 3.4, til 34. Du vil da se at verdien i cellen til medianen endrer seg tilsvarende. (Dette gjelder ikke for tabellen over beskrivende størrelser du beregnet ovenfor). Glem ikke å sette den endrete data-verdien tilbake til den opprinnelige verdien etterpå.

- (iii) *Beregning av gjennomsnitt (mean)*: Skritt for skritt. Først skal vi summere observasjonene. Velg en celle, for eksempel under dataene, B52, og skriv "SUM" i cellen til venstre. Klikk på B52 og deretter på Σ -knappen på verktøylinja. Excel gir et forslag til område det skal summeres over. Hvis dette området ikke stemmer med det du ønsker å summere over, markerer du selv det

ønskete område, eller redigerer formelen som vises i vinduet over arket. Deretter trykker du ENTER eller klikker på haken til venstre for vinduet over arket, og summen kommer fram i B52.

Nå kan vi beregne gjennomsnittet lett, for eksempel i cellen under, B53, ved å klikke på den og deretter skrive formelen: $=B52/49$ og deretter ENTER. Alternativt, klikk først på B53. Skriv så likhetstegnet, = , klikk på cellen B52, og skriv deretter /49 (se hele tiden hvordan formelen kommer fram i vinduet over arket). Trykk til slutt på ENTER.

Når du foretar flere beregninger i et regneark, er det lurt å skrive med ord i arket hva du beregner slik at du husker det neste gang du åpner Excel-filen. Så skriv "Gjennomsnitt" inn i cellen til venstre, A53. Du vil da kanskje se at ordet blir kuttet siden kolonnen er for smal. Dette fikser du ved å markere (klikke) cellen (A53), og deretter fra menyen: *Format -> Column -> AutoFit Selection* som tilpasser kolonnebredden. Du kan også peke på grensen mellom A og B i kolonnehodet og dra ut bredden passende.

Antall desimaler i svaret i B53, kan du regulere ved knappene på verktøylinja som er markert med noen nuller og en vannrett pil. (Husk å klikke på B53 først).

Merk at de to skrittene kunne vært gjort i ett ved å skrive formelen $=SUM(B2:B50)/49$ inn i en celle, deretter trykke ENTER.

En tredje måte å finne gjennomsnittet på, er ved funksjonen, AVERAGE, som du finner på omtrent samme sted som MEDIAN du brukte ovenfor. Prøv det og legg svaret i C53 for å sammenligne med B53.

Sjekk så at svarene i cellene du har laget formler, blir oppdatert når du endrer på en eller flere av de opprinnelige dataene.

- B.** *Lag histogram over dataene.* Excel er skreddersydd for å lage søylediagram av typen fig. 2.2 – fig 2.4 i Løvås, men, dessverre, litt primitiv når det gjelder å lage histogrammer av statistisk art for kontinuerlige data, som beskrevet i Løvås. For eksempel er det vanskelig (*jeg vet ikke hvordan det kan gjøres, men kanskje noen vet om et triks?*) å lage histogrammer med varierende intervall-bredde.

Gå inn i dialogen, *Tools -> Data Analysis -> Histogram*. Legg området for dataene inn i vinduet for Input Range. I neste vindu (Bin Range) kan du angi hvor intervallgrensene du selv har angitt ligger på arket. Hvis du ikke skriver noe her, vil Excel selv gi et forslag. Ellers, velg intervallgrensene selv. La oss si at data ligger mellom 0 og 45. Legg for eksempel inn intervall-grensene, 10, 20, 30, 40, i cellene F2:F5 (hvis de er ledige). Dette gir 5 intervaller: 0 – 10 (med 10 inklusiv), 10 – 20 (inkl. 20), 20 – 30 (inkl.), 30 – 40 (inkl), og større enn 40. Spesifiser deretter F2:F5 (eventuelt ved å markere feltet med musa) i Bin Range – vinduet. (Husk å klikke på vinduet først slik at indikatoren blinker der).

For å få output i samme ark som data, velger du Output Range og spesifiserer i det tilhørende vinduet den første cellen (for eksempel H2) i området der du vil at output skal havne. Output er her hyppighetstabellen bak histogrammet. Nederst i dialogen merker du av Chart Output som gir selve grafen av

histogrammet. Hvis denne er umerket, får du bare hyppighetstabellen, som hos meg ble

| <i>Bin</i> | <i>Frequency</i> |
|------------|------------------|
| 10 | 29 |
| 20 | 13 |
| 30 | 4 |
| 40 | 2 |
| More | 1 |

som viser at 29 observasjoner var større enn 0 og mindre eller lik 10, 13 observasjoner var større enn 10 og mindre eller lik 20, osv.

Hvis du glemte å merke av Chart Output, kan du markere denne tabellen etterpå og bruke Chart Wizard –knappen på verktøy-linja.

Hvis du synes histogrammet ser for flatt eller lite ut, kan du endre det ved å markere diagrammet (klikk på det slik at små firkanter dukker opp på rammen rundt), og så dra i firkantene med pila. Hjørne-firkantene endrer grafen proporsjonalt, men firkantene i midten strekker ut eller presser sammen grafen. Merk at elementene som grafen består av kan redigeres ved å dobbeltklikke på de respektive elementene.

Histogrammet som kommer fram er et søylediagram med mellomrom mellom søylene. Du kan fjerne disse mellomrommene ved å markere søylene (klikk på en av dem slik at små firkanter dukker opp inne i dem. Klikk så to ganger på en av søylene slik at en dialog bestående av flere under-dialoger, kommer fram. Få så fram under-dialogen *Options* og sett 0 i vinduet til *Gap Width*.

Det er nå en av svakheten til Excel-histogrammet kommer fram. Tallene på X-aksen synes å gi et feilaktig bilde av klasse-intervallene (se selv). *Jeg har ikke funnet noen lur måte å fikse dette på, men er åpen for ideer.* Hvis du synes tallene på X-aksen er for misvisende, kan du fjerne dem helt. Marker først X-aksen (klikk på den slik at to små firkanter dukker opp på begynnelsen og slutten av linja). Dobbeltklikk så på linja slik at en redigerings-dialog kommer fram. Få så fram under-dialogen *Pattern* der du merker av *None* under Tick mark labels.

(De fleste statistikk og økonometri-pakker vil vanligvis ha langt mer fleksibel og effektiv grafikk for histogram enn Excel).

Lag også et histogram med intervallstørrelse på 5 (i stedet for 10).

- C. Histogrammet viser en klart skjev fordeling. Ofte kan vi få fram en mer symmetrisk fordeling ved en passende transformasjon av data. Vi skal prøve to transformasjoner. La dataene betegnes med X_1, X_2, \dots, X_n :

- $Y_i = \sqrt{X_i} = X_i^{0.5}$
- $Z_i = \sqrt[4]{X_i} = X_i^{0.25}$

Lag (beregner) en kolonne for Y -ene og en kolonne for Z -ene. Legg for eksempel Y -ene ved siden av X -ene i kolonne C (C2:C50). Skriv Y i C1. Beregn først C2: Klikk på C2 og legg inn formelen $=B2^{0.5}$ og trykk på ENTER eller haken ved formel-vinduet over arket. Klikk på C2 igjen og kopier denne (*Edit -> Copy* fra menyen eller enklere med Ctrl+C). Marker så området der C2 skal kopieres til, fra C3 til C50 og sett inn (*Edit -> Paste* eller enklere Ctrl+V). Merk at du med dette ikke kopierer *tallet* i C2 men derimot formelen bak tallet i C2 (som du ser i vinduet over arket).

Merk at når det ikke står dollartegn i referansen B2, så er referansen relativ, som betyr at den endrer seg ved kopiering/flytting. Når vi for eksempel kopierer celle C2, der formelen $=B2^{0.5}$ står, til celle C3 og C4, vil formelen endres til $=B3^{0.5}$ i C3 og til $=B4^{0.5}$ i C4 osv. Tallene som formlene produserer vil vise seg i cellene, mens formlene som har laget tallene vil vises i vinduet over arket.

Referanser kan gjøres mer eller mindre faste ved dollartegn:

- $\$B2$ betyr at kolonnen (B) holdes fast (ikke endrer seg ved kopiering), mens raden (2) er relativ – dvs. endres ved kopiering til en annen rad.
- $B\$2$ betyr at kolonnen er relativ (endres ved kopiering) og raden er fast.
- $\$B\2 betyr at både kolonne og rad er faste – dvs. endres ikke ved kopiering.

Beregn på samme måten Z -ene i kolonne D. Lag så histogram for både Y -ene og Z -ene. Hvilken av de to transformasjonene synes du ga mest symmetrisk fordeling?