1. Kursusaften

Studerende

Kære Studerende,

Velkommen som studerende i Data, Marked og Afsætning. Dette er din første kursusgang, og det er ment som en blød opstart, hvor du bliver introduceret til fagets indhold og allerede i dag, tilegner dig de første færdigheder i faget. Hav meget gerne i baghovedet, at eksamen består af en mundtlig eksamen (på baggrund af en allerede afleveret synopsis). Øv dig derfor så vidt muligt i at kunne forklare fagets centrale metoder mundtligt.

### Emner på kursusaften 1

Introduktion til faget

Deskriptiv statistik

### Læringsmål

Efter lektionen skal den studerende være i stand til at:

* Gennemføre beskrivende statistiske undersøgelser af et talmateriale, herunder
  + hyppigheder og frekvenser (evt. kumulerede hyppigheder og kumulerede frekvenser)
  + pindediagrammer og histogrammer
  + bestemmelse af gennemsnit, varians og spredning for både diskrete og grupperede observationssæt
* Undersøge, om et givent statistisk talmateriale er normalfordelt (normalfraktildiagram)
* Gennemføre en statistisk beskrivelse af data fra et finansielt marked
* Have kendskab til standardafvigelsen som risikomål for afkastet for finansielle aktiver

### Forberedelse (bør laves forinden)

Litteratur:

[ST]

[2 Datasæt og data](https://s.tepedu.dk/datast-og-data.html)

[3 Deskriptiv statistik](https://s.tepedu.dk/deskriptiv-statistik.html)

Hjemmeopgaver: Ingen

### Opgaver og aktiviteter til lektionen:

*Bemærk, at alt talmateriale til nedenstående opgaver findes i vedlagte Excel-ark.*

Som noget af det første bør man have hente hjælpeværktøjet Freestat fra onlinebogen.

**Introduktion til Excel**

**Vi kommer til at anvende Excel i faget, og det er vigtigt, at du forstår hvordan programmet grundlæggende fungerer.Der ligger videoer og øvelser i onlinebogen .**

**Opgave 1**

Selandia Bank gennemførte i april 2017 en kampagne for at skaffe flere kunder. For at kunne vurdere kampagnens effektivitet har man registreret det daglige antal af nye kunder i måneden før kampagnen, dvs. marts 2017, og i måneden efter kampagnen, dvs. maj 2017.

I marts 2017 var der 23 arbejdsdage, og antallet af nye kunder pr. dag var:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | 4 | 0 | 4 | 4 | 3 | 2 | 5 |
| 4 | 7 | 0 | 2 | 3 | 1 | 0 | 7 |
| 2 | 4 | 1 | 5 | 1 | 1 | 2 |  |

I maj 2017 var der 21 arbejdsdage, og antallet af nye kunder pr. dag var:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 8 | 4 | 3 | 4 | 3 | 4 | 7 |
| 2 | 6 | 6 | 0 | 6 | 3 | 11 |
| 6 | 7 | 9 | 6 | 1 | 7 | 6 |

a) Hvorfor er det nødvendigt at tage tallene for marts 2017, altså fra før kampagnen, i betragtning for at kunne vurdere kampagnens effektivitet?

b) Hvis man faktisk konstaterer, at der i gennemsnit er flere nye kunder om dagen i maj end i marts, kan man så konkludere, at det er reklamekampagnen, der er årsag til dette?

c) Beregn i Excel gennemsnittet af det daglige antal nye kunder i marts 2017.

Beregn endvidere det tilsvarende gennemsnit for maj 2017.

Det er af stor vigtigt at kende til fordelingen af antallet af nye kunder og ikke blot gennemsnittene. For at kunne finde denne fordeling skal man først finde hyppighederne, altså finde antallet af dage med 0 nye kunder, med 1 ny kunde etc. Disse optællinger kan man foretage manuelt, men ved større datasæt er det bedre at få Excel til det.

Man kan benytte pivottabeller i Excel til at sortere data. I kapitel 2. i bogen [Metode Data for finansøkonomer](https://excel1.tepedu.dk/), kan du se videoer og opgaver der viser hvordan pivottabeller fungerer.

Vi har nu hyppighederne for det første observationssæt, marts 2017. Det ses, at der således var 5 dage, hvor der ankom 4 nye kunder.

d) Find hyppighedstabellerne for hver af de to observationssæt, marts 2017 og maj 2017. (Fordi der er to forskellige observationssæt, så skal der også være to forskellige tabeller. Den ene er vist ovenfor. Bemærk, at man ikke må blande de to datasæt sammen og lave en fælles hyppighedstabel).

Hyppighedstabellerne er i sig selv ikke så interessante, og som regel ønsker man at illustrere frekvenserne (de relative hyppigheder) grafisk. Da der er tale om diskrete observationer (antallet af nye kunder er altid et helt tal), så er et pinde- eller stolpediagram en god måde af illustrere fordelingen af data på.

e) Lav de to pinde- eller søjlediagrammer for de to observationssæt.

**Opgave 2**

Et af statistikkens mange anvendelsesområder er indenfor analyse af aktier og andre værdipapirer. Følgende opgave løfter lidt af sløret for disse anvendelser:

Når man overvejer køb af en given aktie, skal man ikke blot tage afkastet i betragtning, men også risikoen. Det er jo ikke en god ide at investere i en aktie med meget høj risiko, hvis man kan finde en med samme afkast, men lavere risiko. Men hvad menes der egentligt med risiko?

Inden vi kan besvare dette spørgsmål, så skal vi finde nogle aktiedata at kigge på. På Københavns Fondsbørs hjemmeside http://www.nasdaqomxnordic.com, kan man finde aktuelle og historiske kurser for en lang række aktier og andre værdipapirer.

Lad os fokusere på to givne aktier, nemlig Carlsberg B og Novo Nordisk. Data for kalenderåret 2010 er gemt i det vedlagte regneark.

Her er der mange tal, men de interessante er lukkekurserne dag for dag, som findes i søjlen Closing Price. Disse lukkekurser er placeret på et separat faneblad i Excel-arket.

a) Lav en graf i Excel, som viser begge aktiers kursudvikling dag for dag gennem året 2010.

Umiddelbart ser de to aktier ud til at have klaret sig nogenlunde lige godt gennem året, omend kursen for Novo Nordisk er vokset lidt mere end for Carlsberg.

I virkeligheden er sådanne kurser ikke så interessante i sig selv, men man skal se på kursændringerne og specielt på de procentvise kursændringer:

Betragt to (fiktive) aktier X og Y. 1. januar 2010 koster aktie X 100 kr., mens aktie Y koster 1.000 kr. Et år senere koster de to aktier henholdsvis 200 og 1.200 kr. Hvilken aktie var den bedste investering?

Beregner man naivt forskellene, så vil man på en X-aktie have tjent 200-100 = 100 kr., mens man på en Y-aktie ville have tjent 1.200 – 1.000 = 200 kr., altså det dobbelte. Så Y-aktien var den bedste investering! - eller hvad?

Regner man fortjenesten ud i procent af investeringsbeløbet, så fås et andet billede:

X-aktien: Y-aktien:

X-aktien er altså den bedste investering. Havde man 1.000 kr. til sin rådighed, så kunne man have købt en Y-aktie og tjent 200 kr., eller købt 10 X-aktier og tjent 10 gang 100 kr., dvs. 1.000 kr.

Man skal derfor betragte afkastet på aktien. Vi vil derfor i det følgende beregne det daglige afkast for hver af de to aktier, Carlsberg B og Novo Nordisk.

b) Beregn det daglige afkast for hver af de to aktier.

(Vink: Man kan naturligvis ikke beregne det daglige afkast for den første dag, 4. januar. For 5. januar og Carlsberg kan man anvende formlen =(B3-B2)/B2)

(Bemærk, at vi i denne beregningsmåde ikke tager hensyn til eventuelle udbetalte dividender på de to aktier. Dividendeudbetalinger sker heldigvis så sjældent, at det ikke får nogen større praktisk betydning for resultatet).

c) Beregn det gennemsnitlige daglige afkast for de to aktier.

Fordi de daglige afkast kan antage alle mulige værdier og ikke bare hele tal, så bør de grupperes.

d) Optæl hyppighederne for afkastene for hver af de to aktier. Anvend Excel pivottabeller anvend en inddeling med grænserne -8%, -7%, -6%, ..., 0%, 1%, 2%, ..., 9%, 10%.

Bemærk opbygningen af Excels output. Når der udfor eksempelvis -4% står tallet 2, så betyder det, at der er 2 observationer i intervallet mellem -5% og -4%, dvs. to observationer, som er større end -5%, og som er mindre end eller lig med -4%.

e) Lav histogrammer for data fra opgave d.

f) Deskriptiv statistik, Grupperede observationer beregner også gennemsnittet af observationerne. Ser man efter, så er dette beregnede gennemsnit ikke det samme tal, som man fik i delopgave b ovenfor. Hvorfor? Hvad er mest rigtigt?

**Opgave 3**

Ser man på de to histogrammer fra opgave 2e, så ligner de mere eller mindre hinanden – histogrammerne er 'klokkeformede' eller 'tilnærmelsesvist normalfordelte', som det hedder i fagsproget.

Det er en vigtig egenskab, for har man et datasæt, som er tilnærmet normalfordelt, kan man anvende en række statistiske teknikker, som vi skal se nærmere på senere i kurset.

Desværre er det i praksis ret svært at identificere normalfordelinger ud fra histogrammer og pindediagrammer, så vi anvender et trick: Vi tegner et normalfraktildiagram for datasættet. I normalfraktildiagrammet repræsenteres hver observation af et punkt, og ligger disse punkter mere eller mindre langs en ret linje, så er der tale om en normalfordeling.

a) Undersøg ved hjælp af normalfraktildiagrammer, om afkastene for de to aktier fra opgave 3 er normalfordelte.

Normalfordelingen optræder rigtig mange steder, så det er en særdeles anvendelig model.

b) Undersøg ved hjælp af normalfraktildiagrammer, om observationssættene fra opgave 1 er normalfordelte.

**Opgave 4**

Vi har i opgave 2 set på to aktier og beregnet deres gennemsnitlige daglige afkast. Afkastet fortæller jo, om man kan tjene penge på at investere i aktien, men afkastet fortæller ikke hele sandheden – man skal også se på risikoen.

Ét risikomål, som ofte anvendes ved aktieinvesteringer, er den såkaldte standardafvigelse eller spredning. Dette begreb måler sandsynligheden for at få et afkast, som afviger væsentligt fra det gennemsnitlige afkast. Generelt gælder, at jo større standardafvigelse af afkastet, jo større er sandsynligheden for at opnå et lille eller endog negativt afkast, dvs. jo større risiko.

a) Bestem standardafvigelserne af de daglige afkast for de to aktier.

Generelt gælder, at når man har med normalfordelte observationssæt at gøre, så måler standardafvigelsen 'bredden af klokken'.

b) Passer dette i tilfældet med de to aktier?

Man kan naturligvis beregne standardafvigelser af alle mulige observationssæt.

c) Bestem standardafvigelserne for de to observationssæt fra opgave 1.