Påliteligheten til en stikkprøve

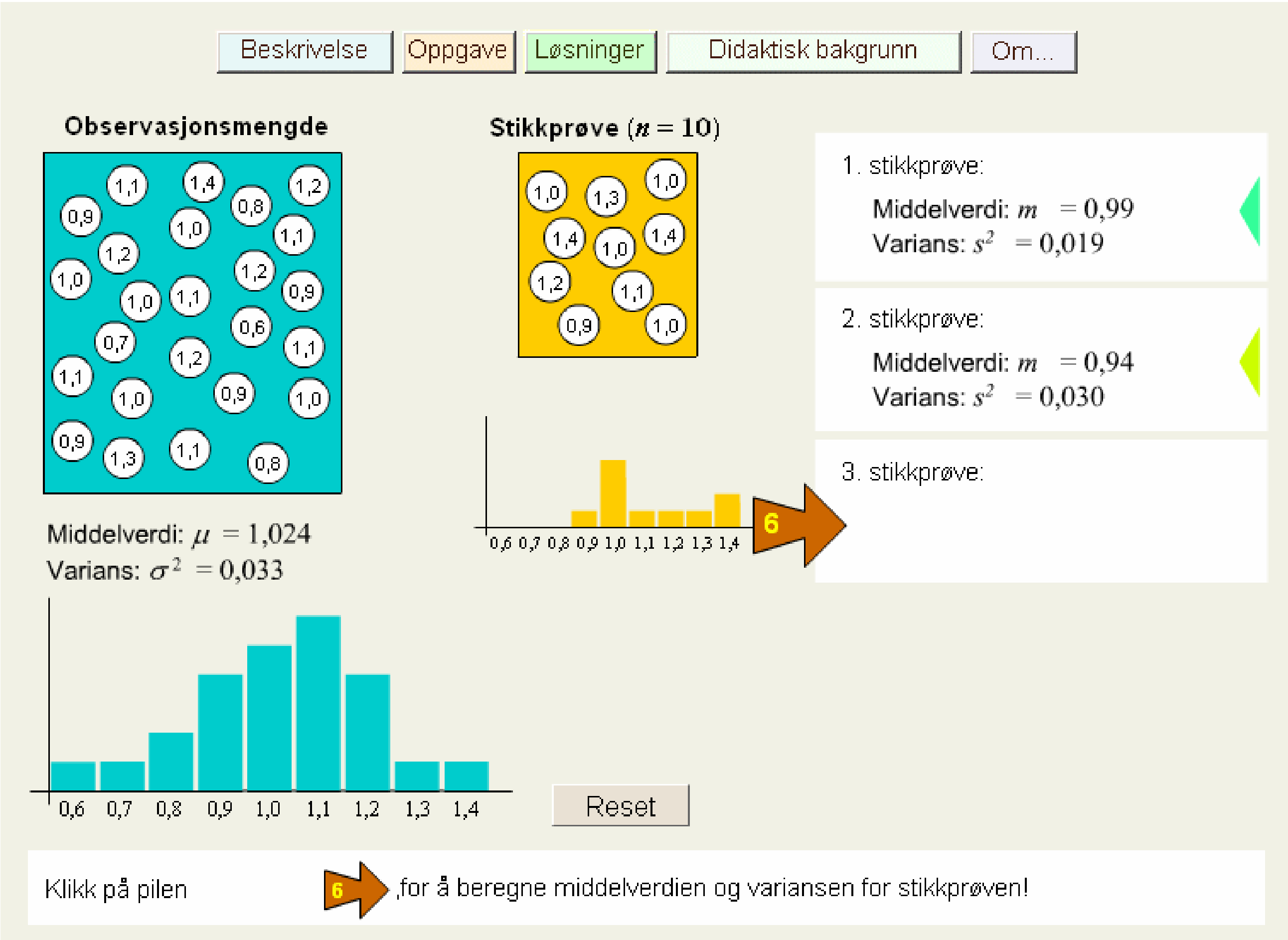
Om originalen

"Zuverlässigkeit einer Stichprobe" på http://www.mathe-online.at/galerie/wstat2/stichprobe/Denne læringsstøtten ble utviklet av Franz Embacher, ved Fakultetet for naturvitenskap og matematikk og Instituttet for teoretisk fysikk ved Universitetet i Wien, og Petra Oberhuemer i forbindelse med prosjektet

"Nye medier i matematikk-utdanningen"

med støtte fra departementet i Østerrike for utdanning, vitenskap og kultur. Universitetet i Wien, mai 2004.

Siden er høsten 2004 oversatt av Cornelia Brodahl, Høgskolen i Agder, og Heidi Mæsel Oftedahl, Kvadraturen Skolesenter, med vennlig tillatelse av forfatterne. Den ligger på Galleri DigiVitalis under Høgskolen i Agder, Fakultet for Realfag.



Beskrivelse

Temaet i denne animasjonen er påliteligheten av de to statistiske parametrene middelverdi og varians som en beregner på grunnlag av en *stikkprøve*. Animasjonens målsetning er å illustrere noen grunnleggende problemstillinger.

I denne forbindelsen forutsettes det, at begrepene middelverdi og standardavvik allerede er kjent og at en kjenner til hva disse begrepene forteller om en observasjonsmengde. Variansen er kvadratet av standardavviket.

Utgangspunktet er en **observasjonsmengde** (karakterisert ved en mengde av mulige utfall, som inntreffer med en kjent sannsynlighet), som man vil undersøke. Så skal denne observasjonsmengde (som det ofte er tilfelle i praksis) ikke være tilgjengelig til direkte statistisk analyse (f.eks. fordi den er for stor). I slike tilfeller må man greie seg med informasjon fra en stikkprøve. I animasjonen har vi en observasjonsmengde der de ulike utfall og sannsynligheten som de opptrer med karakterisert gjennom følgende tabell:

Utfallet	opptrer med en sannsynlighet på
0,6	0,04
0,7	0,04
0,8	0,08
0,9	0,16
1,0	0,20
1,1	0,24
1,2	0,16
1,3	0,04
1,4	0,04

De viktigste statistiske parametrene for denne observasjonsmengden er (avrundet):

- Middelverdi: $\mu = 1{,}024$
- Standardavvik: $\sigma = 0{,}1817$
- Varians (kvadrat av standardavviket): $\sigma^2 = 0{,}033$

Det blir tatt en stikkprøve på 10 utfall ut av vår observasjonsmengde (i henhold til ovenfor angitte sannsynligheter) og det beregnes *deres* middelverdi *m* og varians s^2.

Spørsmål: Med hvilken sikkerhet kan en trekke konklusjoner for de ulike parametrene til observasjonsmengden?

For å kunne svare på dette spørsmålet, forestiller vi oss (i et tankeeksperiment), at *flere* stikkprøver blir tatt:

- **Trinn 1 - 6**:
  I denne animasjonen blir det vist tre stikkprøver. De faller alle litt ulikt ut: Middelverdi (*m*) og variansen (s^2) i hvert av de tre tilfellene, avviker noe fra hverandre. Dette forteller noe om at man må regne med en **statistisk usikkerhet**. Et mål for denne usikkerheten får vi når vi undersøker, hvordan disse stikkprøvene skiller seg fra hverandre. Dette vil vi utføre i de påfølgende trinn. (Ved hjelp av trekantene som står til høyre for dataene kan en studere stikkprøven og deres hyppighetstabell også under senere trinn.)

- **Trinn 7**:
  Så beregnes middelverdien til de tre stikkprøve-middelverdiene. For informasjon om hvordan denne parameteren forholder seg for et *stort tall* av stikkprøver, kan du bevege musen over den røde teksten "Les meg 1"!
  NB: Symbolet E(...) som blir brukt i teksten betegner middelverdien (gjennomsnittet) av et stort antall stikkprøver (og betegnes som forventningsverdien).

- **Trinn 8**:
  Variansen til de tre stikkprøve-middelverdiene beregnes også. For mer informasjon om hvordan denne parameteren forholder seg for et *stort tall* av stikkprøver, kan du bevege

musen over den blå teksten "Les meg 2 "!
NB: Standardavviket (spredning) Δm for stikkprøve-middelverdiene for et stort antall stikkprøver er definert ved formelen $\Delta m^2 = E((m - E(m))^2)$ som også kan uttrykkes slik: $\Delta m^2 = E(m^2) - E(m)^2$.

- **Trinn 9**:
  Ut av de tre stikkprøve-variansene beregnes middelverdien. For informasjon om hvordan denne parameteren forholder seg for et *stort tall* av stikkprøver, kan du bevege musen over den grønne teksten "Les meg 3"!

Animasjonen inneholder ingen utledning av formlene som presenteres i lesmeg-tekstene, men skal illustrere deres betydning.

Ved hjelp av "Reset"-knappen kan du til en hver tid vende tilbake til begynnelsen av animasjonen. De tre viste stikkprøvene er faste og blir brukt på nytt ved ny omstart.

Hva en kan lære ut av dette:

Prøv ved hjelp av animasjonen (og formlene i de tre "Les meg" -tekstene, som du helst skal notere ned) å bli bevisst på følgende sammenheng:

Selv om verdiene til de statistiske parametrene m og s i en stikkprøve *ikke* står fast i utgangspunktet, kan en angi "forventninger" (så fremt observasjonsmengden er kjent). Dette blir brukt til å undersøke det i praksis viktige tilfelle, at observasjonsmengden er *ukjent*.

Etter en har tatt en stikkprøve, kan en gjøre følgende anslag ut av kjennskapet til m og s:

- Det beste anslag for den (ukjente) middelverdien μ for observasjonsmengden er middelverdien m av stikkprøven.
- Den statistiske usikkerheten (spredning) Δm av dette anslaget kan en også angi: Spredningen beregner en ut av standardavviket s og antall stikkprøver n

$$\Delta m = \frac{\sigma}{\sqrt{n}}$$

Denne formelen betegnes også iblant som "roten av n - regelen ". Denne uttrykker den innlysende setningen "Jo større stikkprøven, desto mer nøyaktig kan utsagnene om observasjonsmengden bli som vi kan utlede av denne." Mer presist: For å *halvere* sikkerheten i anslaget av observasjonsmengdens middelverdi, må en *firedoble* omfang av stikkprøven! Men legg merke til at standardavviket σ på høyre side er en ukjent størrelse.

- Det beste anslaget for standardavviket σ til observasjonsmengden er den såkalte "**empiriske spredning** " (også kalt "**midlere feil til en enkeltmåling**")

$$s \cdot \sqrt{\frac{n}{n-1}}$$

For store n er den tilnærmet lik standardavviket.

- Setter man de to resultatene sammen, får en et anslag for den statistiske usikkerheten.

$$\Delta m \approx \frac{s}{\sqrt{n-1}}$$

Parameteren på den høyre siden blir også betegnet som "**midlere feil for middelverdiene**". Formelen tillater å angi påliteligheten til slutningen på observasjonsmengden" ene og alene ut ifra utfallene i stikkprøven. For store n har vi den enkle formelen

$$\Delta m \approx \frac{s}{\sqrt{n}}$$

- Alle disse formlene er kun gyldige i sammenhenger der utfallene i en stikkprøve er uavhengig av hverandre ("med tilbakelegging"). Stikkprøven kan da ha et vilkårlig omfang, dvs.n kan bli vilkårlig stor. I praksis tas stikkprøver derimot ofte "uten tilbakelegging" ut av en gitt grunnmengde (f.eks. i meningsmålinger, når hver person maksimalt blir spurt én gang). Blir sannsynligheten for å trekke ut et element av observasjonsmengden flere ganger så liten at en kan se bort i fra den, kan en bruke formelen over også for dette tilfellet. (Dersom dette ikke er tilfellet, så er påliteligheten av anslaget på grunnlag av en stikkprøve til og med mindre! I et grensetilfelle kan stikkprøven innbefatte hele observasjonsmengden, da forsvinner all statistisk usikkerhet.)

Et konkret **eksempel** på dette blir behandlet i oppgave 5.

Oppgaver

1. Hvilken av de tre stikkprøvene i animasjonen beskriver observasjonsmengden best?
2. Beregn størrelsen Δm^2 (dvs. gjennomsnittsvariansen ut fra et *stort* antall stikkprøver, se "Les Meg 2"-teksten og **Beskrivelse**) og Δm for situasjonen som er behandlet i animasjonen!
3. Hvilke av middelverdiene av de tre stikkprøvene ligger lengre fra middelverdien til observasjonsmengden enn Δm ?
4. Hvor sannsynlig er det at en stikkprøve har middelverdien lik 0,6?
5. En skal finne gjennomsnittlig kroppshøyde til befolkningen i et land. I den forbindelse blir det tatt en stikkprøve på 100 (tilfeldig valgte) personer. Standardavviket til de målte kroppshøydene er 40 cm. Ut fra de 100 målte kroppshøydene blir middelverdien beregnet. Hvor nøyaktig angir denne midlere kroppshøyde til befolkningen?

   Hvordan endrer den statistiske usikkerheten seg, når en i stedet tar en stikkprøve på 10000 personer, når standardavviket fortsatt er 40 cm?
6. Betrakt observasjonsmengden som bare har to elementer, -1 og 1, som hver opptrer med en sannsynlighet på 1/2! Dette tilsvarer et myntkast, der resultat mynt gir poengverdi 1 og resultat krone gir poengverdi -1. Vis at forventningsverdien for observasjonsmengden er μ = 0 og standardavviket for observasjonsmengden er σ = 1! Når du kaster mynten n ganger og adderer poengene - hva forventer du at summen vil bli?

Løsningsforslag

1. I forhold til middelverdi kommer den første stikkprøven nærmest observasjonsmengden, sett i forhold til standardavvik (eller varians) kommer den andre og tredje stikkprøven best ut. Spørsmålet har altså ingen entydig løsning.
2. Avrundet er $\Delta m^2 = 0.0033$ og $\Delta m = 0.057$.
3. Den andre og tredje.
4. Siden 0,6 er den minste av alle mulige resultatverdier, får man en stikkprøve med middelverdi 0,6 bare når en trekker denne verdien ti ganger. Sannsynligheten til å trekke denne verdien én gang er 1/25. Sannsynligheten til å trekke denne verdien 10 ganger er $(1/25)^{10} \approx 10^{-14}$.
5. Tips: Bruk formler som er angitt i LesMeg-tekstene (og i "Beskrivelse" til animasjonen) med

   n = 100
   s = 40 cm.

   Da viser det seg at den statistiske usikkerheten knyttet til beregningen av middelverdi for kroppshøyden, er ca 4 cm. (Den eneste nødvendige "utregningen" er $100^{-1/2} = 1/10$).

   Det betyr: Den virkelige gjennomsnittlige kroppshøyden av befolkningen kan avvike med flere centimeter fra middelverdien som er beregnet på grunnlag av en stikkprøve.

Ta man en stikkprøve med 10000 personer, og standardavviket fortsatt er 40 cm, reduseres den statistiske usikkerheten til omlag 4 mm (siden $10000^{-1/2} = 1/100$).

6. Formelen i "Les Meg 2"-teksten reduseres da i dette tilfelle til $\Delta m = n^{-1/2}$.

   Formelen beskriver det forventede avviket mellom stikkprøvens middelverdi for poengsummen og den virkelige middelverdi. Siden den virkelige middelverdien er 0, kommer stikkprøvenes middelpoengverdi til å ligge i størrelsesorden på (pluss eller minus) $n^{-1/2}$, men blir på den andre siden lik summen til de kastede poengverdiene delt på *n*. Den forventede summen ligger derfor antageligvis i en størrelsesorden på (pluss eller minus) $n^{1/2}$.

   Hvis du altså kaster mynt en million ganger (et "stort" tall), så ligger den forventede middelverdien på (pluss eller minus) 1/1000 (et "lite" tall) og den forventede summen ligger ved (pluss eller minus) 1000 (et "mellomstort" tall). Påliteligheten til stikkprøven ved et anslag for middelsverdien til observasjonsmengden er i dette tilfellet allerede ganske godt. Summen av poengverdiene, "netto-overskudd" på (pluss eller minus) 1000, er i forhold til antall myntekast (1 million) lite.

   Dette resultatet har to interessante anvendelser:

   - I et lykkespill kan du i hver omgang tape eller vinne 1 krone med samme sannsynlighet. Når du begynner med en startkapital på 0 kroner og spiller n omganger, vil saldoen din (ved stor *n*) antagelig ligge i størrelsesorden (pluss eller minus) $n^{1/2}$. Det vil si at etter mange omganger har du enten tapt eller vunnet mye. *Hvilken* av disse to situasjonene som inntreffer kan en naturligvis ikke forutsi. Når det først har oppstått et tap, blir sannsynligheten for at du kan utligne det allerede oppståtte tapet stadig mindre!
   - Den Brownske molekularbevegelsen eller "tilfeldige gange": Hvis du ser på en liten støvpartikkel i en vanndråpe i et mikroskop, vil du se at den beveger seg hit og dit i et uregelmessig mønster - det ser nesten ut som om den lever! Dette heter Brownsk bevegelse etter mannen som beskrev fenomenet først. Blir en partikkel i hver tidsenhet utsatt for sammenstøt med molekyler og forskjøvet i en lengde L til høyre eller venstre (til en hver tid med samme sannsynlighet), vil den etter n støt ha tilbakelagt en strekning i størrelsesorden $L\, n^{1/2}$ - om det var til venstre eller til høyre lar seg ikke forutsi.

Didaktisk bakgrunn

En analyse av påliteligheten til en stikkprøve er vanskelig, framfor alt fordi her har vi en kombinasjon av to tilfeldighetsprosesser:

- en stikkprøve at et omfang *n* ("med tilbakelegging") og
- dette tilfeldige utvalget er *ett* av alle mulige stikkprøver.

Allerede betegnelsen "midlere feil for middelverdiene" som er en viktig størrelse i statistikk gjør det språklig tydelig at den dobbelte tilfeldighetsstrukturen er komplisert.

Animasjon prøver å skape forståelse for hva som er ment med dette og liknende begrep og hva de formlene som er brukt står for. Animasjonen kan brukes helt uavhengig av om en skal utvikle formelen algebraisk eller ikke. Det anbefales

- å gå gjennom animasjonen flere ganger, slik at den underliggende logikken blir klar
- og å skrive ned formlene som en finner i "Les Meg"-tekstene
- å lese teksten som ligger under knappen "Beskrivelse" nøye.

At det nødvendige regnearbeidet (en trenger stort sett bare å trekke kvadratroten) er forholdsvis lite må ikke skjule at den situasjon som vi betrakter er kompleks. Et godt mål for oppnådd forståelse er evnen til ikke bare å kunne utføre beregninger, men også å kunne gjengi resultatene noenlunde språklig korrekt.