# Læringsmål i STAT100

Kathrine Frey Frøslie

16 9 2021

## STAT100: Et innføringskurs i 7 moduler (13 uker)

STAT100 er et 10-stp innføringskurs som er delt opp i 7 moduler:

M1 Deskriptiv statistikk

M2 Sannsynlighetsteori

M3 Estimering

M4 Hypotesetesting

M5 Bivariate analyser

M6 ANOVA

M7 Regresjon

Læringsutbyttebeskrivelsene (i kurset brukes ordet "læringsmål" for å gjøre det lettere for de som leser dårlig) for hver av modulene beskriver konkret hva studentene skal kunne, og er formulert så de er verifiserbare, både for studenter og undervisere (Ref Felder & Brent, Teaching and learning in STEM).

Lista er et godt utgangspunkt for arbeidet med å få en samstemt undervisning, det som på engelsk kalles "constructive alignment". I dette ligger at vi skal tilstrebe et meningsbærende samsvar mellom *hva* studentene skal lære (læringsutbyttebeskrivelsene, i STAT100 kalt læringsmål), *hvordan* de skal lære det (læringsaktiviteter) og hvordan vi *kan vite* at de faktisk har lært det de skal (vurderingsformer).

Eksempelvis vil godt formulerte læringsmål være til hjelp for

- å vurdere hvilke deler av læreboka som er relevant
- å vurdere hvilke oppgaver som er relevante/nødvendige
- å se hvilke temaer som er dekket i dette kuset, så videregående kurs kan bygge på det (constructive alignment på tvers av kurs)
- å vurdere om kurset kan erstattes av andre kurs, eller om dette kurset kan erstatte andre kurs
- å diskutere innholdet i kurset med andre, både internt og mellom utdanningsinstitusjoner

Resten av dette dokumentet lister opp læringsmålene i kurset slik de er formulert på den datoen dokumentet viser.

## M1 Deskriptiv statistikk (1 uke)

Når modul 1 er ferdig (og de to første fysiske samlingene er over), skal du ha lært grunnleggende begreper og prinsipper for hvordan du oppsummerer et datasett og leser oppsummeringer som andre har laget.

Mer spesifikt skal du kunne

1. Forklare hva et datasett og hva en variabel i et datasett er, og identifisere hvilke typer variabler et gitt datasett består av.
2. Definere de observerbare størrelsene relativ frekvens, gjennomsnitt, standardavvik, median og kvartiler, og forklare hvilke egenskaper ved en variabel (i et datasett) de beskriver.
3. Forklare i hvilke situasjoner vi velger å oppsummere variabler vha frekvenstabeller, gjennomsnitt og standardavvik, og i hvilke situasjoner vi i stedet velger median og kvartiler, og hvorfor.
4. Skissere fordelingen til en variabel ut i fra deskriptive mål.
5. Gjøre en kritisk vurdering av den deskriptiv statistikken som oppgis i en vitenskapelig artikkel eller et medieoppslag
6. Gjøre deskriptiv analyse vha RStudio

## M2 Sannsynlighetsteori

Når modul 2 er ferdig, skal du ha lært et utvalg av grunnleggende begreper og prinsipper innen sannsynlighetsteori.

Mer spesifikt, skal du kunne følgende om sannsynlighetsregning (lærebokas kapittel 3):

1. Forklare hva som menes med stokastiske (tilfeldige) forsøk, utfall, utfallsrom og hendelser (begivenheter), og identifisere disse i gitte situasjoner
2. Definere sannsynlighet ut i fra relativ frekvens (De store talls lov) og bruke dette til å beregne sannsynligheter i gitte situasjoner, samt gjøre en vurdering av om n er stor nok til at den relative frekvensen er tilstrekkelig lik sannsynliheten. (De siste to punktene overlapper med estimering, se modul 3.)
3. Definere union, snitt og komplement av begivenheter, hva som menes med disjunkte begivenheter, og regne ut sannsynligheter basert på slike opplysninger.
4. Definere betingede sannsynligheter og beregne slike i gitte situasjoner.
5. Forklare forskjellen på disjunkte begivenheter og uavhengige begivenheter.
6. Definere uavhengighet og bruke regneregler til å avgjøre om begivenheter er uavhengige, samt sette opp (og utlede) Bayes regel og bruke den til å beregne betingede sannsynligheter.
7. Beskrive en uniform sannsynlighetsmodell og definere sannsynlighet ut i fra gunstige/mulige, og bruke dette til å beregne sannsynligheter i gitte situasjoner.

Du skal kunne følgende om stokastiske variabler (lærebokas kapittel 4):

1. Definere «Stokastisk variabel», avgjøre om en gitt stokastisk variabel er diskret (tellevariabel) eller kontinuerlig (målevariabel), og sette opp verdimengden for variabelen.
2. Beskrive forskjellen på en sannsynlighetsfordeling for en diskret stokastisk variabel, og en sannsynlighetstetthet for en kontinuerlig stokastisk variabel, forklare hvordan sannsynligheter beregnes for de to typene, og hvorfor kumulativ sannsynlighet er nyttig for å beregne sannsynligheter.
3. Beskrive de vanligste oppsummeringstallene for sannsynlighetsfordelinger, og hvilke egenskaper ved fordelingene de oppsummerer.
4. Definere forventningsverdi, varians og standardavvik for en diskret og en kontinuerlig stokastisk variabel, og regne ut forventningsverdi, varians og standardavvik for en diskret variabel ut i fra verdimengden og sannsynlighetsfordelingen til variabelen.
5. Bruke regneregler for forventningsverdi til å regne ut forventningsverdi til lineærtransformasjoner av stokastiske variabler (både avhengige og uavhengige variabler).

6. Bruke regneregler for varians til å beregne variansen til en lineærkombinasjon av uavhengige stokastiske variabler, og deretter beregne standardavviket til den samme transformasjonen.
7. Forklare kort hvordan man kan finne median og kvartiler for stokastiske variabler.
8. Definere kovarians og korrelasjon og være i stand til å regne seg fra den ene størrelsen til den andre.
9. Bruke regneregler for varians og kovarians til å beregne variansen til en sum eller differanse av to avhengige stokastiske variabler, og deretter beregne standardavviket til den samme transformasjonen.

Om binomisk fordeling (lærebokas kapittel 5.2) skal du kunne:

1. Gjengi betingelsene som må være oppfylt for at en stokastisk variabel $X$ skal være binomisk fordelt
2. Identifisere situasjoner der betingelsene for en binomisk kan sies å gjelde, og der en binomisk sannsynlighetsmodell vil være en god modell
3. Kunne beregne forventningsverdi, varians og standardavvik for en binomisk fordeling ut i fra modellparameterene
4. Kunne beregne binomiske punktsannsynligheter og kumulative sannsynligheter ved hjelp av formelen for den binomiske sannsynlighetsfordelingen, og ved å bruke tabellen med kumulative binomiske sannsynligheter (Tabell E.1 i læreboka).
5. Forklare prinsippet bak normalfordelingstilnærmingen til den binomiske fordelingen, fortelle når den er relevant, og hvilken normalfordeling som ligner mest på en gitt binomisk fordeling.
6. Beregne tilnærmede binomiske sannsynligheter ved hjelp av normalfordelingstilnærmingen

Og om normalfordelingen (lærebokas kapittel 5.7) skal du kunne:

1. Beskrive kjennetegnene til en normalfordeling, gi eksempler på stokastiske variabler som kan antas å være normalfordelte, og oppgi forventningsverdi, varians og standardavvik for en normalfordeling ut i fra modellparameterene
2. Vite forskjellen på en generell normalfordeling og en standard normalfordeling
3. Kunne standardisere en generell, normalfordelt stokastisk variabel og bruke normalfordelingstabell (Tabell E.3 i læreboka) til å beregne sannsynligheter i en hvilken som helst normalfordeling
4. Bør kjenne til noen utvalgte sannsynligheter og tilhørende kvantiler og (sprednings)intervaller i en normalfordeling
5. Forklare hvordan sannsynligheter og kvantiler i normalfordelingen kan brukes til å tolke deskriptiv statistikk.
6. Avgjøre i hvilke situasjoner normalfordelingen kan sies å være en god modell for det som er observert, blant annet: Kunne identifisere normalitet og avvik fra normalitet ved hjelp av histogrammer og normalfordelingsplott (ofte kalt kvantil-kvantil-plott, QQ-plott eller qq-plott)
7. Kunne forklare prinsippet bak sentralgrenseteoremet og bruke det til å finne fordelingen til et gjennomsnitt av mange observasjoner, samt gi en uformell vurdering av om antallet observasjoner kan sies å være tilstrekkelig til at sentralgrenseteoremet kan brukes
8. Finne og formulere fordelingen til en sum av uavhengige, normalfordelte variabler.
9. Forklare hva som er forskjellen på en standard normalfordeling og en $T$-fordeling, og i hvilke situasjoner disse fordelingene er omtrent like.

## M3 Estimering

Når modul 3 er ferdig, skal du ha lært grunnleggende begreper og prinsipper om estimering, og du skal kunne Forklare prinsippet bak estimering ved å bruke begrepene populasjon og utvalg, parameter og observasjoner, estimator og estimat, beskrive og generalisere.

Mer spesifikt, skal du kunne

1. Forklare forskjellen på deskriptiv statistikk og estimering.

2. Utlede og forklare hvorfor gjennomsnittet er en forventningsrett estimator for $\mu$ og en observert andel (en relativ frekvens) er en forventningsrett estimator for $p$.
3. Forklare forskjellen på punktestimat og konfidensintervall.
4. Forklare forskjellen på standardavviket til observasjonene og estimeringsusikkerheten til estimatoren, utlede standardavviket til estimatorer for $\mu$ og for $p$, beregne den observerte standardfeilen (det estimerte standardavviket; den observerte estimeringsusikkerheten) til disse estimatorene og forklare hvilken rolle estimeringsusikkerheten spiller i estimeringsprosessen.
5. Sette opp fordelingen til estimatoren for $\mu$ (den stokastiske variabelen gjennomsnittet) i fire situasjoner:
   - Når vi antar normalfordelte data, kjent $\sigma$ ($N$-fordeling)
   - Når vi antar ikke-normalfordelte data, kjent $\sigma$ ($\approx N$-fordeling)
   - Når vi antar normalfordelte data, ukjent $\sigma$ (T-fordeling)
   - Når vi antar ikke-normalfordelte data, ukjent $\sigma$ ($\approx N$-fordeling),
     og forklare hvilket prinsippet som brukes til å finne fordelingen i hver situasjon
6. Bruke en $T$-fordelingstabell
7. Sette opp fordelingen til estimatoren for $p$, basert på normalfordelingstilnærmingen til en binomisk fordeling.
8. Utlede, regne ut, tolke og vurdere den praktiske betydningen av:
   - Et estimat og tilhørende konfidensintervall for en forventningsverdi ($\mu$) og
   - Et estimat og tilhørende konfidensintervall for en andel i en populasjon ($p$).
9. Vurdere hvilke alternativer vi har for å estimere $\mu$ når vi har ikke-normalfordelte data, ukjent $\sigma$ og liten $n$.
10. Gjøre en utvalgsberegning; regne ut hvor stor $n$ vi må ha for at et konfidensintervall (med en gitt konfidensgrad)

- for $\mu$ ikke skal være bredere enn en oppgitt maksimalbredde
- for $p$ ikke skal være bredere enn en oppgitt maksimalbredde

## M4 Hypotesetesting

Når modul 4 er ferdig, skal du ha lært grunnleggende begreper og prinsipper i statistisk hypotesetesting.

Mer spesifikt skal du kunne

1. Forklare hva hypotesetesting er i det statistiske fagfeltet, og gi et historisk eksempel på betydningen av hypotesetesting.
2. Definere og gi eksempler på nullhypotese og alternativ hypotese, ensidig og tosidig alternativ.
3. Definere Type I-feil og Type II-feil og forklare hvorfor den ene feilen utelukker den andre, og hvorfor vi ikke kan minimere sannsynligheten for begge disse feilene samtidig.
4. Definere og forklare hva som menes med signifikansnivået $\alpha$, og vurdere hva som er et riktig nivå for en hypotesetest.
5. Forklare hva som menes med en testobservator, og gi eksempler på testobservatorer.
6. Formulere fordelingene til testobservatorer for $\mu$ og testobservator for $p$ under forskjellige antakelser (om observasjonene antas normalfordelt eller ikke-normalfordelt $\pounds X$ med kjent eller ukjent $\sigma$ og liten eller stor $n$, eller binomisk fordelt med liten eller stor $n$).
7. Definere forkastningsgrense og forkastningsområde for en testobservator.
8. Definere hva $p$-verdien for en gitt observasjon er.
9. Forklare forskjellen på å konkludere ut i fra et forkastningsområde og å konkludere ut i fra en $p$-verdi.
10. Sette opp korrekte hypoteser (med ensidig eller tosidig alternativ) for en gitt situasjon og utføre en hypotesetest med et gitt signifikansnivå for $p$ eller for $\mu$, både ved

- å beregne forkastningsområdet og sammenligne observasjonene med dette, og ved

- å beregne $p$-verdien for en gitt observasjon, og sammenligne den med $\alpha$.

11. Tolke sammenhengen mellom hypotesetesting og konfidensintervall

## M5 Bivariate analyser

Når modul 5 er ferdig, skal du ha lært å utføre hypotesetester i alle situasjoner der du er interessert i å undersøke sammenhengen mellom to variabler, der variablene er kategoriske eller kontinuerlige.

Mer spesifikt skal du kunne bruke bivariate analyser på riktig måte og i riktig situasjon, herunder

1. Forklare hva som er felles for korrelasjonsanalyse, to-utvalgs t-test, enveis ANOVA og Pearsons kjikvadrat-test, og gjøre rede for prinsippene bak hver av analysemetodene.
2. Beregne en korrelasjonskoeffisient vha RStudio og gi en praktisk tolkning av hva størrelsen betyr, samt å sette opp hypoteser og utføre en test for om det er signifikant sammenheng mellom de to variablene.
3. Utføre en to-utvalgs t-test (herunder være i stand til å beregne både forkastningsområde, p-verdi og et konfidensintervall for differansen), både for hånd og vha RStudio, inkludert å sette opp hypoteser, formulere modell, testobservator, modellantakelser, samt vurdere om antakelsene er oppfylt, trekke riktig konklusjon og tolke resultatene i en praktisk situasjon.
4. Utføre en kjikvadrattest for sammenhengen mellom de to kategoriske variablene i en krysstabell (herunder være i stand til å beregne både forkastningsområde og p-verdi), både for hånd og vha RStudio, inkludert å sette opp hypoteser, formulere modell, testobservator, modellantakelser, beregne forventede frekvenser i tabellen, samt vurdere om antakelsene er oppfylt, trekke riktig konklusjon og tolke resultatene i en praktisk situasjon.

## M6 ANOVA

Når de to ukene i modul 6 er ferdig, skal du ha lært grunnleggende begreper og prinsipper for enveis variansanalyse (ANalysis Of VAriance, ANOVA).

Mer spesifikt skal du kunne

1. Gjenkjenne og identifisere situasjoner der man har et datasett som består av to eller flere grupper med uavhengige (kontinuerlige) målinger, og der målet er å sammenligne gruppene for å avgjøre om de er like, eller om noen grupper skiller seg ut, som situasjoner som er aktuelle å analysere vha enveis ANOVA.
2. Formulere en sannsynlighetsmodell for en enveis ANOVA, inkludert å identifisere og tolke variabler og parametere i modellen, og å formulere hvilke antakelser som må være oppfylt for at analysen skal være fornuftig å gjøre.
3. Vurdere ut i fra dataene som er tilgjengelig, om antakelsene for modellen er oppfylt.
4. Forklare analyseprinsippene ved enveis ANOVA, herunder dekomponeringen av variasjonen.
5. Regne ut de ulike kvadratsummene for enkle eksempeldatasett.
6. Formulere nullhypotese og alternativ hypotese for en gitt ANOVA, både i form av modellparametere, og med ord som beskriver den faktiske situasjonen.
7. Regne ut en F-testobservator fra oppgitte kvadratsummer i et eksempel (oftest en utskrift fra RStudio), og være i stand til å identifisere fordelingen til denne testobservatoren under nullhypotesen.
8. Bruke RStudio-utskrifter til å finne estimater for alle parameterene i modellen.
9. Skal kunne identifisere og forklare de ulike tallene i en ANOVA-utskrift fra RStudio, og bruke dem til å avgjøre om nullhypotesen skal beholdes eller forkastes, for et gitt signifikansnivå.
10. Skal kunne formulere en passende kontrast for å gjøre spesifikke gruppesammenligninger i en situasjon der nullhypotesen i en enveis ANOVA forkastes. Må også kunne formulere hypoteser for kontrasten, både formulert ved parametere, og med ord.
11. Skal kunne finne et estimat og den tilhørende standardfeilen til en gitt kontrast, vite hvilken fordeling testobservatoren for kontrasten har, og beregne en p-verdi for kontrasten i en gitt situasjon.

## M7 Regresjon

Når de to ukene i modul 7 er ferdig, skal du ha lært grunnleggende begreper og prinsipper for lineær regresjonsanalyse med én forklaringsvariabel, såkalt "enkel lineær regresjon".

Mer spesifikt skal du kunne

1. Gjenkjenne og identifisere situasjoner der man har et datasett som består av to (eller flere) variabler, der den ene variabelen antas å være en responsvariabel og består av uavhengige (kontinuerlige) målinger, og en annen variabel i datasettet (eller flere andre variabler i datasettet), kan brukes til å predikere eller forklare noe av variasjonen i responsvariabelen som situasjoner som er aktuelle å analysere vha regresjonsanalyse.
2. Forklare at i enkel regresjonsanalyse vil det være én (kontinuerlig) responsvariabel og én prediktor eller forklaringsvariabel (oftest kontinuerlig eller binær).
3. Formulere en sannsynlighetsmodell for en enkel regresjonsanalyse, inkludert å identifisere og tolke variabler og parametere i modellen, og å formulere hvilke antakelser som må være oppfylt for at analysen skal være fornuftig å gjøre.
4. Gjøre en kritisk vurdering av om antakelsene for modellen er oppfylt, ut i fra opplysningene som er oppgitt og dataene som er tilgjengelig,.
5. Forklare analyseprinsippene ved enkel regresjonsanalyse, herunder minste kvadraters metode.
6. Forklare hva som menes med predikerte verdier og residualer, og hva de kan brukes til.
7. Forklare hva som er forskjellen på å bruke regresjonsanalyse til prediksjonsformål og til estimeringsformål, forskjellen på bruk og tolkning av konfidensintervall og prediksjonsintervall, og velge riktig hovedstrategi i en gitt analysesituasjon.
8. Formulere nullhypotese og alternativ hypotese for regresjonsparametere for en gitt situasjon, også med ord som beskriver den faktiske situasjonen.
9. Sette opp testobservatorer for regresjonsparameterene og fordelingene til testobservatorene under nullhypotesene.
10. Bruke RStudio-utskrifter til å finne estimater for alle parameterene i modellen, og til å beregne både konfidensintervaller og prediksjonsintervaller for regresjonsparameterene når det er aktuelt.
11. Tolke estimater fra en RStudio-utskrift, og bruke utskriften til å avgjøre om nullhypotesene skal beholdes eller forkastes, for et gitt signifikansnivå.
12. Gjøre deskriptiv analyse vha RStudio