STK 1000 høsten 2003: Oblig I

Dette er oppgavesettet for første obligatoriske innleveringsprosjekt for kurset STK 1000. Det leveres ut **onsdag 17/9/3**, direkte fra forelesningen, og fra resepsjonskontoret ved Matematisk institutt, 7. etasje i Abel-bygget. Leveringsfrist er **mandag 29/9/3**, enten til instituttkontoret ved Matematisk institutt (i så fall senest kl. 15:00 den mandagen), til din gruppelærer, eller direkte til foreleser (ettersom det er forelesning i kurset, i Sophus Lies auditorium, kl. 14:15 – 16:00 den dagen).

Hvis flere samarbeider om å løse oppgavene, må likevel hver student levere sin besvarelse; spesielt kreves det at hver student har samlet sitt eget datamateriale for Oppgave 1. Det må gå frem av besvarelsen hvem du eventuelt har samarbeidet med.

Du kan levere håndskrevet eller maskinskrevet (tekstbehandlet) besvarelse. Øverst til høyre på første side skriver du (tydelig!) ditt navn, samt nummer på den gruppe du er knyttet til (eventuelt lærerens navn). Der du bruker MINITAB må utskrifter og plott legges ved, eller klebes inn i besvarelsen. Du vil få bruk for introduksjonsheftet *Starthjelp i* MINITAB*: Versjon 13 for Windows*, som kan skaffes fra Kabul eller Akademika. Spesielt er det instruksjoner om utskrift fra MINITAB i dette heftets seksjon 10.

Softwarepakken MINITAB skal brukes mer utstrakt siden i kurset, for eksempel i forbindelse med Oblig II (med leveringsfrist mandag 10/11/3). Ved denne anledning kan man klare seg nesten uten MINITAB. Oppgave 1 kan gjerne gjøres «helt manuelt», inkludert histogrammene og beregning av gjennomsnitt og standardavvik, men de som vil kan her gjerne anvende MINITAB. Oppgave 2 er essensielt en øvelse om normalfordelingen og bruk av bokens Table A. Oppgave 3 i dette settet tjener det enkle formål at dere skal tvinges inn i MINITAB (eventuelt for første gang), produsere noe forholdsvis banalt, og så berge dere helskinnet ut igjen, medbringende et bevis på at dere faktisk har vært der.

En besvarelse vil kunne bli godkjent selv om man ikke har svart perfekt på hvert delspørsmål. Jeg bestemmer meg herved også for at en besvarelse som er meget god m.h.t. Oppgave 1 og 2 vil bli godkjent, selv om Oppgave 3 ikke er besvart. For Oblig II vil det imidlertid bli avkrevet MINITAB-bruk og do. kunnskaper for å få besvarelsen godkjent.

Oppgave 1

HVOR FORSKJELLIGE ER ENGELSK OG NORSK, med hensyn til ordlengde, preposisjonshyppighet, tegnsettingsiver, setningskonstruksjoner, leddsetningsmønstre, osv.? Kan vi se forskjell på Sigrid Undset og Jane Austen, ut fra slike rent kvantitative mål?

Vi skal her bry oss om ett enkelt av disse aspektene, nemlig ordlengdene. Gå til din bokhylle, og velg én bok på norsk og én på engelsk. Dette kan være romaner, novellesamlinger, prosa-artikler eller noe helt annet, men skal ikke være teknisk fagstoff, med for eksempel matematiske formler. For hver av disse to bøkene skal du så gå gjennom følgende øvelse.

Slå opp på side 101, eventuelt på første «normale tekstside» etter side 101 i det tilfelle at denne siden ikke er en vanlig tekstside. Gå så gjennom hvert ord på siden, og noter antall bokstaver i ordet.

(a) Oppgi hvilke to bøker du har valgt (gi forfatter, tittel, forlag, årstall).

(b) Sammenfatt ordlengderesultatene i to tabeller (en for hver av de to bøkene), og i to histogrammer (en for hver bok).

(c) Kall disse ordlengdene $x_1, x_2, \ldots, x_n$, der x_1 er antall bokstaver i ord nr. 1, x_2 er antall bokstaver i ord nr. 2, osv., og n er antall ord på den side du har undersøkt. Angi n for den norske og den engelske side. Beregn videre gjennsomsnittet $\bar{x}$ og medianen M for dine to datasett.

(d) Beregn dessuten standardavviket s for de to datasettene.

(e) Kommenter eventuelle likheter og forskjeller mellom de to datasettene.

(f) Det er forelesers intensjon at alle enkeltstudentenes datasett skal tas vare på i en liten database, som så kan analyseres videre, for eksempel mandag 1. desember. Er det problemstillinger du mener kan være av spesiell interesse, som kunne forfølges og belyses, gjennom et slikt datamateriale? Med andre ord, er det gode spørsmål du mener kursets foreleser bør stille på eksamen i dette kurset?

Oppgave 2

Lengdehoppere hopper imponerende langt, men naturligvis ikke like langt hver gang. Vi skal se på de to fremragende hoppere Semper Excelsior og Dr. Livingstone I. Presume. Presume er stabil, på ganske høyt nivå, mens Excelsior har lavere gjennomsnitt, men større utslag fra hopp til hopp. Konkret skal vi anta at Presumes hopp kommer fra en $\mathrm{N}(790, 15)$-fordeling, altså en normalfordeling med senter (teoretisk forventning) 790 cm og spredning (teoretisk standardavvik) 15 cm, mens Excelsiors hopp analogt oppfører seg om tall trukket fra en $\mathrm{N}(780, 20)$-fordeling. Her regner vi kun med «normale» lengdesprang, og ser for eksempel bort fra de såkalte døde hopp.

Her blir du bedt om å beregne ulike sannsynligheter. Kjære størrelser har mange navn, og språkbruken rundt «sannsynligheter» er noe fleksibel. Når vi studerer sannsynligheten for at Presume skal hoppe lenger enn 8.10 meter, for eksempel, er dette det samme som å beregne hyppigheten (i prosent, eller på skalaen fra 0 til 1) av hans fremtidige hopp som vil være lenger enn 810 cm.

(a) Finn sannsynlighetene p_E og p_P for at henholdsvis Excelsior og Presume i et gitt sprang skal hoppe kortere enn 785 cm.

(b) Finn tilsvarende sannsynlighetene for at henholdsvis Excelsior og Presume i et gitt hopp skal kommer lengre enn 8 meter.

(c) I løpet av Excelsiors karriere har han naturligvis noen hopp som er gloriøst lange, og noen som er sørgelig korte. Hvor lange er Excelsiors fem prosents lengste sprang? Og hvor korte er Presumes fem prosents korteste hopp?

(d) Hva er sjansen for at Presume skal vinne over Excelsior, når de får ett hopp hver? – For å svare på dette trenger du følgende resultat, som vi skal lære mer om siden i kurset; se spesifikt Chapter 4.4 i boken. Resultatet det siktes til er at hvis X er $\mathrm{N}(\mu_1, \sigma_1)$ og Y er $\mathrm{N}(\mu_2, \sigma_2)$, og disse stokastiske variablene er uavhengige av hverandre, så er differansen $X - Y$ også normalfordelt, med senter $\mu_1 - \mu_2$ og spredning $\sqrt{\sigma_1^2 + \sigma_2^2}$.

Oppgave 3

Som nevnt i innledningen til oppgavesettet er denne oppgave 3 ikke ment som noe mer enn en øvelse for å «komme inn i» MINITAB. Senere i kurset skal vi naturligvis anvende denne softwarepakken til mer intelligente formål.

Oppgaven du skal utføre er som følger. Den krever at du har tilgang til MINITAB, enten via det brukernavn og passord du skal ha fått tildelt qua kursdeltager, eller på annen måte.

Finn en terminal, logg deg inn, og finn frem til MINITAB; konsulter om nødvendig innføringsheftet. Du kan også be din gruppelærer om hjep.

Gå til rullegardinsmenyen og gå via `Calc` til `Random Data` og `Uniform`. Ved hjelp av dette skal du generere 100 uniformt fordelte tilfeldige tall i kolonnen C11. Gratulerer. Så skal du, ved hjelp av `Calc` og `Column Statistics` beregne gjennomsnitt (`Mean`) og median av de 100 tallene. Rapporter disse to sentermålene.

Så skal du produsere et histogram over de 100 tallene du har generert. Dette gjøres via `Graph` og `Histogram`. Histogrammmet skal så printes ut og legges ved besvarelsen; du kan igjen konsultere startheftet, seksjon 10, eller spørre din gruppelærer.

Nils Lid Hjort