STK 1000 høsten 2003: Oblig II

Dette er oppgavesettet for det andre obligatoriske innleveringsprosjekt for kurset STK 1000. Det leveres ut **onsdag 29/10/3**, direkte fra forelesningen, og fra resepsjonskontoret ved Matematisk institutt, 7. etasje i Abel-bygget, og vil også bli lagt ut på kursets hjemmeside. Leveringsfrist er **mandag 10/11/3**, 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 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.

En besvarelse vil kunne bli godkjent selv om man ikke har svart perfekt på hvert delspørsmål.

Oppgave 1

HVA ER KONFIDENS? Tenk deg at du har 60 tilfeldige utvalg, hver av størrelse 30, fra en $N(10, 3)$-fordelt populasjon (altså en normalfordeling med forventning 10 og standardavvik 3). For hvert av de 60 utvalgene kan du finne et 90% konfidensintervall for forventningsverdien (som vi her vet er lik 10) ved å bruke metodikk og formler fra lærebokens Chapter 6.1 (se f.eks. side 422).

(a) Hvor mange av de 60 konfidensintervallene venter du vil inneholde den riktige verdien, altså $\mu = 10$?

(b) La Y være en stokastisk variabel som angir hvor mange av de 60 konfidensintervallene som vil inneholde den riktige verdien 10. Hvilken fordeling har Y? Begrunn ditt svar.

(c) Hva er forventning og standardavvik for Y?

– Du kan bruke MINITAB til å simulere den situasjonen som er beskrevet innledningsvis i oppgaven. Først trekker du 60×30 stokastiske variable som er uavhengige og $N(10, 3)$-fordelte. Dette gjør du ved å gå til `Calc`, så til `Random Data`, og så til `Normal`. Generer her 30 rader med data og ta vare på dem i kolonnene `C1-C60`. Tallene i en kolonne gir deg nå de 30 observasjonene i et tilfeldig utvalg, mens de 60 kolonnene gir deg de 60 utvalgene.

  Nå skal du finne et konfidensintervall for hvert av de 60 utvalgene. Kommandoen er fra `Stat` til `Basic Statistics` og `1-Sample Z`. Husk at konfidensgraden skal være 90%.

(d) Utfør kommandoene ovenfor og angi de 60 konfidensintervallene du får. Hvor mange av dem inneholder den riktige forventningen 10? Kommenter resultatet i lys av punkt (b).

Oppgave 2

SKØYTELØPERE BLIR OGSÅ SLITNE, når de bare har løpt langt nok. Likevel har mange skøyteløpere høyere gjennomsnittsfart på stayerdistansen 1500 meter (tre og tre-kvart runde) enn på sprintdistansen 500 meter (en og en-kvart runde). Den firedobbelte verdensmester og olympiamester Ådne Søndrål, for eksempel, har 1.45.26 på 1500 m (Salt Lake City 2002), som svarer til en gjennomsnittlig 500 m-fart på 105.26/3 = 35.087, betydelig raskere enn hans beste 500 m-tid 35.72 (Hamar 2000). Mange andre løpere har dog høyere fart på en 500 m enn på en 1500 m. For å gripe litt bakover i tiden kan man ihukomme Knut «Kuppern» Johannesen, mannen hvis verdensrekordtid på 10.000 m er nevnt i minst syv norske skjønnlitterære romaner. Hans 2.09.4 (fra Karuizawa 1963) er en del langsommere (43.133 pr. 500 m) enn hans 42.3 (fra Squaw Valley 1960).

Vi skal i denne oppgaven undersøke hvor vanlig det er at løpere har høyere tempo på 1500 m enn på 500 m, og skal også finne ut om det er en signifikant forskjell, eller om forskjellene i tempo kan regnes som tilfeldige variasjoner. Til dette formål konsulterer jeg den såkalte Adelskalenderen, over verdens beste skøyteløpere gjennom tidene, basert på deres personlige rekorder og deres sammenlagtpoengsum (som er $x_1 + x_2/3 + x_3/10 + x_4/20$, der x_1, x_2, x_3, x_4 er tidene på de fire klassiske distanser 500 m, 1500 m, 5000 m, 10000 m). For dagens formål plukker jeg ut de 32 beste (der sampelstørrelsen er valgt av nasjonalromantiske grunner, idet de fire beste norske for tiden er plassert som nr. 25, 28, 29, 32). For disse 32 helter beregner jeg brøkene

$$R = \frac{(x_2/3)}{x_1},$$

altså 500-fart under 1500-løpet, delt på 500-tiden. Observasjonene $R_1, \ldots, R_{32}$ er som følger:

```
0.954  0.966  0.978  0.973  0.983  0.976  0.984  0.968  0.947  0.970
0.953  0.964  0.984  0.966  0.976  0.998  0.980  0.968  1.010  0.968
0.988  0.985  0.995  1.007  0.956  0.978  0.975  0.993  0.971  0.984
0.966  0.982
```

(a) Lag et histogram for de 32 ratio-variablene. Beregn også gjennomsnitt, median og standardavvik.

(b) La μ betegne den underliggende middelverdi for hele R-populasjonen, altså R-observasjonenes teoretiske forventningsverdi. Finn et konfidensintervall for μ med konfidensgrad 99%, idet du antar at R-ene kommer fra en populasjon med et standardavvik på 0.017. Gjør rede for hvilke antagelser du gjør for å kunne tolke intervallet som et 99% konfidensintervall.

(c) Vi er altså interessert i om R-ratioenes gjennomsnitt i populasjonen av gode skøyteløpere er lik 1, eller ikke. Sett opp en test for hypotesen $H_0\colon \mu = 1$, mot alternativet $H_a\colon \mu \neq 1$, idet du fortsatt antar at standardavviket for enkeltratioene er $\sigma = 0.017$. Utfør testen, og beregn (eller anslå) dessuten P-verdien. Formuler en konklusjon.

(d) Hva kan det komme av, at mange skøyteløpere altså har høyere gjennomsnittstempo på tre og tre-kvart runde, enn på en og en-kvart runde?

Oppgave 3

Det er mulig at utvalget av skøyteløpere gjort over ikke er helt representativt, når det gjelder problemstillingen om 1500 m løpes i høyere tempo enn 500 m. For å undersøke dette nærmere plukker jeg i tillegg ut 32 noe mindre glamorøse løpere, nemlig de som i øyeblikket befinner seg på plassene fra 269 til 300 på Adelskalenderen. Så beregner jeg R-brøken også for disse, og får

1.008	0.985	1.023	0.967	0.992	0.973	1.026	0.984	0.984	0.988
0.986	1.002	0.993	0.984	1.002	0.999	0.988	1.001	0.972	0.997
0.985	0.979	1.003	1.007	1.040	0.951	1.006	0.978	1.037	0.967
0.996	0.978								

Disse ses igjen på som uavhengige observasjoner med et underliggende populasjonsmiddeltall μ'.

Her er det to naturlige problemstillinger: (i) Gjelder det samme funn som for Oppgave 2, med hensyn til hypotesen om at μ' er lik 1? (ii) Det er naturligvis store forskjeller mellom nivået til den supergruppen jeg samplet fra først (de 32 beste i verden) til den noe mer pedestriøse gruppen jeg studerer i denne oppgaven (fra nr. 269 til nr. 300 på Adelskalenderen). Men er det forskjell med hensyn til ratioen mellom 1500 m-fart og 500 m-fart?

(a) Test, som i forrige oppgave, om $\mu' = 1$ eller ikke, og formuler en konklusjon.

(b) Test hypotesen at de to gruppene av skøyteløpere er like, med hensyn til ratioen mellom 1500-fart og 500-fart. Gjør rede for de antagelser du bruker, og formuler en konklusjon.

Men no skrid Kuppern i Squaw Valley!
Eg har ikkje tenkt å gå nokon 10000 m,
men ordi fær ein djervare sving,
og ho mor grip fastare um staven.

– Olav H. Hauge

Nils Lid Hjort