Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

Faglig kontakt under eksamen:

Ingelin Steinsland	926 63 096
Ola Diserud	932 18 823
Arvid Næss	995 38 350

TMA4245 Statistikk

Lørdag 26. mai 2012 kl. 9–13

Hjelpemidler: Gult A5-ark med egne håndskrevne notater (stemplet av Institutt for matematiske fag), *Tabeller og formler i statistikk* (Tapir forlag), *Matematisk formelsamling* (K. Rottmann), kalkulator HP 30s eller Citizen SR-270X

Sensur: 18. juni 2012

Oppgave 1 Snødensitet

Densiteten (massetettheten) av snø er høyst 1 kg/dm^3 (veldig våt snø). Anta at sannsynlighetstettheten for densiteten i kg/dm^3 av en tilfeldig valgt snøprøve er gitt ved $f(x) = \beta(\beta+1)x(1-x)^{\beta-1}$, $0 \le x \le 1$, der β er en positiv parameter.

a) Anta, bare i dette punktet, at $\beta = 2$.

Hva er sannsynligheten for at en tilfeldig valgt snøprøve har densitet mellom 0,5 og 0,9 kg/dm^3?

b) Finn sannsynlighetsmaksimeringsestimatoren for β basert på et tilfeldig utvalg X_1, X_2, $\ldots$, X_n av snødensiteter.

Hva blir estimatet hvis $n = 100$ og $\sum_{i=1}^{n} \ln(1-x_i) = -104{,}0$?

Oppgave 2 Temperatur i mars og april

I år (2012) var det mange steder i Norge kaldere i april enn i mars.

La X være gjennomsnittstemperaturen i mars og Y gjennomsnittstemperaturen i april ved Værnes et tilfeldig valgt år, begge målt i °C. Anta at X er normalfordelt med forventningsverdi μ_m og varians σ^2, og at Y er normalfordelt med forventningsverdi μ_a og varians σ^2.

a) Anta at vi har data i form av verdier av X og Y for et tilfeldig utvalg av år. Forklar hvordan du grafisk kan undersøke om antakelsen om normalfordeling holder. Hvordan kan du grafisk undersøke om X og Y er uavhengig?

Skisser også grovt hvordan slike grafiske framstillinger kan se ut både når antakelsen om normalfordeling er oppfylt og når den ikke er det, samt når X og Y er uavhengig og når de ikke er det.

Gjennomsnittstemperaturen i °C ved Værnes for årene 2001–2012 var slik:

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
x_i (mars)	−2,5	0,5	3,3	2,6	−0,7	−4,6	3,3	0,8	1,9	−0,5	1,2	3,8
y_i (april)	4,1	7,2	5,0	7,9	5,8	4,9	5,0	5,9	6,9	4,8	6,7	3,2

Det oppgis at $\sum_{i=1}^{12} x_i = 9{,}10$, $\sum_{i=1}^{12} y_i = 67{,}40$, $\sum_{i=1}^{12} x_i^2 = 77{,}07$, $\sum_{i=1}^{12} y_i^2 = 399{,}30$, $\sum_{i=1}^{12}(x_i - y_i)^2 = 364{,}53$, der $i = 1$ står for år 2001, $i = 2$ for 2002 osv.

b) Anta at mars-temperaturene fra år til år er uavhengige. Finn et 99 %-konfidensintervall for forventet gjennomsnittstemperatur i mars.

Vi ønsker ved hypotesetesting å prøve å påvise at differansen mellom forventet gjennomsnittstemperatur i april og mars er mindre enn 5 °C. Anta at også april-temperaturene er uavhengige fra år til år.

c) Sett opp hypotesene.

Vi kan enten bruke en toutvalgstest eller en partest. Hva vil du gjøre? Argumenter for valget ditt, og utfør testen du velger. Bruk signifikansnivå $\alpha = 0{,}05$.

Oppgave 3 Kjemisk fabrikk

En maskin på en fabrikk utfører en prosedyre for å lage et kjemikalium. Et giftig biprodukt dannes i en mengde på X gram hver gang maskinen utfører prosedyren, der X er normalfordelt med forventningsverdi 20 og standardavvik 4. Hvis mer enn 25 g av biproduktet dannes, lyser ei varsellampe til prosedyren er avsluttet. Maskinen kan innstilles slik at den utfører prosedyren flere ganger på rad, men den kan ikke stoppes før alt er ferdig. Mengden biprodukt er uavhengig fra gang til gang.

a) Hva er sannsynligheten for at lampa lyser når maskinen utfører prosedyren én gang?

Maskinen utfører prosedyren tre ganger. Hva er sannsynligheten for at lampa lyser minst én gang?

Maskinen utfører prosedyren 100 ganger. Finn en tilnærmet sannsynlighet for at lampa lyser 15 eller flere ganger.

b) Forurensingsmyndighetene krever at sannsynligheten for at maskinen produserer 500 g eller mer av biproduktet på en dag, skal være 0,01 eller mindre. Hvor mange ganger kan maskinen utføre prosedyren i løpet av en dag for at dette skal være oppfylt?

Oppgave 4 Grillfest i kveld?

En gjeng studenter vil ta seg en velfortjent kveld fri fra eksamensforberedelser etter statistikkeksamen. De ønsker å grille, men med værforbehold: Det skal verken regne eller være for kaldt (under 15 °C).

a) La A være hendelsen at det regner, B hendelsen at det er kaldt (under 15 °C) og C hendelsen at det er grillvær (15 °C eller mer og ikke regn).

Framstill hendelsene i et Venn-diagram.

Videre vet vi at $P(A) = P(B) = 0{,}4$, og $P(A \cap B) = 0{,}2$.

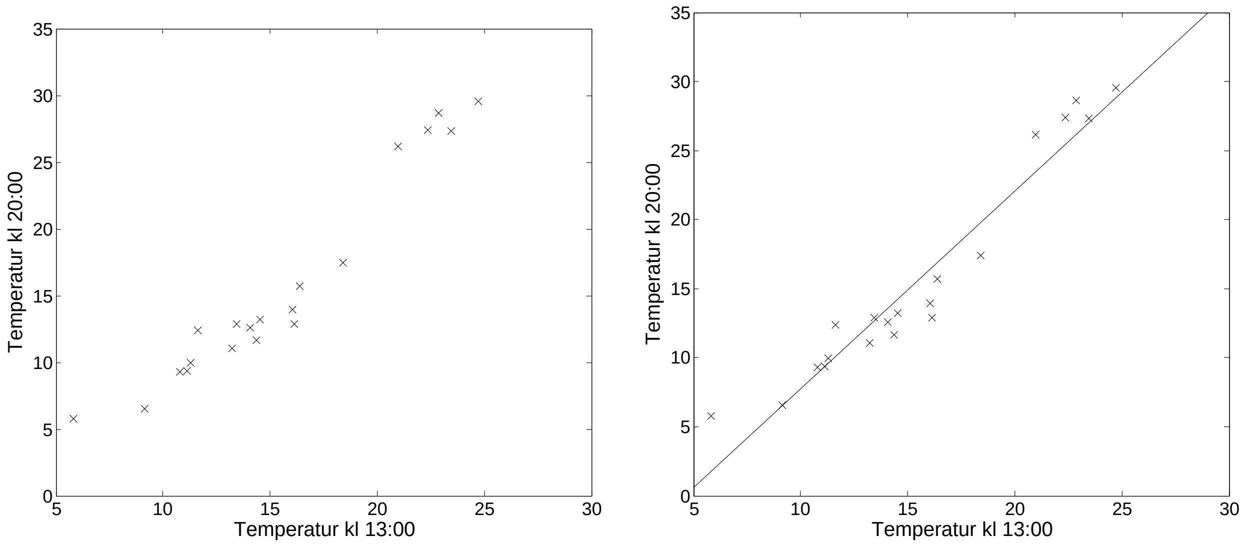
Er A og B disjunkte? Er A og B uavhengige?

Finn $P(C)$.

Er A og C disjunkte? Finn $P(C \mid A')$.

(Begrunn svarene kort.)

De utsetter avgjørelsen til rett etter eksamen, klokka 13, og vil vurdere om det blir grillfest basert på været da. For å gjøre et best mulig valg, lager de en enkel lineær regresjonsmodell med temperatur klokka 13, x, som forklaringsvariabel (uavhengig variabel) og temperatur klokka 20, Y, som responsvariabel (avhengig variabel), begge målt i °C: $Y = \alpha + \beta x + \epsilon$, der ϵ er normalfordelt med forventningsverdi 0 og varians σ^2.



Figur 1: Temperatur kl. 13 og kl. 20 på 26. mai i årene 1992–2011. Estimert regresjonslinje er tegnet inn til høyre.

La x_i være temperaturen klokka 13 og y_i temperaturen klokka 20 på 26. mai i årene 1992–2011, der $i = 1$ står for år 1992, $i = 2$ for 1993 osv. ($n = 20$ år). Hybelverten til en av studentene har data tilgjengelig (figur 1, venstre). Vi antar at dataene er et tilfeldig utvalg fra den nevnte regresjonsmodellen.

La $\hat{\alpha}$ og $\hat{\beta}$ være estimatorene av α og β som en får ved minste kvadraters metode. I punktet nedenfor kan du uten bevis bruke estimatorene oppgitt i *Tabeller og formler i statistikk*, og at $\hat{\beta}$ er forventningsrett, normalfordelt, har varians $\sigma^2/\sum_{i=1}^n (x_i - \bar{x})^2$ og er uavhengig av $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$.

b) Finn sannsynlighetsfordeling, forventningsverdi og varians til $\hat{\alpha}$. Begrunn kort svarene og overgangene i utregningene.

List opp antakelsene som er gjort i regresjonsmodellen. Vurder om mulig fra figur 1 om disse er oppfylt.

Estimatene basert på dataene blir $\hat{\alpha} = -6{,}57$, $\hat{\beta} = 1{,}43$ og $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = 2{,}04^2$. Videre er $\bar{x} = 15{,}5$ og $\sum_{i=1}^n (x_i - \bar{x})^2 = 510{,}7$. Den estimerte regresjonslinja er tegnet inn i figur 1 (høyre).

c) Finn et estimat av forventet temperatur klokka 20 dersom temperaturen klokka 13 er 15 °C.

Finn et 95 %-prediksjonsintervall for temperaturen klokka 20 dersom temperaturen klokka 13 er 15 °C.