

# UNIVERSITETET I OSLO

## Matematisk Institutt

EKSAMEN I:                   **STK 1000 – Innføring i anvendt statistikk**  
  **Avslutningseksamen**  
TID FOR EKSAMEN:       **Mandag 1. desember 2003, kl. 9:00–12:00**  
HJELPEMIDLER:           **Læreboken, kalkulator**

**Dette oppgavesettet utgjør den andre av kursets to eksamener. Det inneholder seks oppgaver og er på fire sider.**

### Oppgave 1

MAN ER INTERESSERT I sannsynligheten  $p$  for en bestemt begivenhet  $A$ . La oss tenke oss at det ikke er mulig å beregne denne  $p = P(A)$  eksakt, men at det er praktisk mulig å utføre uavhengige enkeltforsøk, der  $A$  hver gang opptrer med sannsynlighet  $p$  (og opptrer ikke, med sannsynlighet  $1 - p$ ). La  $X_n$  være antallet  $A$ -forekomster i løpet av de  $n$  første enkeltforsøk, og la  $\hat{p}_n = X_n/n$  være den relative hyppighet for  $A$ .

- (a) Hvilken sannsynlighetsfordeling har  $X_n$ ? Sett opp formler for forventningsverdi og varians for  $X_n$ .
- (b) Man vil estimere  $p$  såpass presist at standardavviket for estimatet  $\hat{p}_n$  er mindre enn eller lik 0.04. Finn ut hvor mange enkeltforsøk man (minst) må utføre for å oppnå dette,
  - (i) dersom det er kjent på forhånd at  $p \leq 0.25$ ;
  - (ii) dersom intet er kjent på forhånd om  $p$ 's verdiområde.
- (c) Anta at man observerte  $X_{300} = 58$  forekomster av  $A$  i løpet av  $n = 300$  enkeltforsøk. Beregn et konfidensintervall for  $p$ , med konfidensgrad tilnærmet 95%, basert på dette.

### Oppgave 2

EN SPEIELL SANNSYNLIGHETSFORDELING er gitt ved funksjonen

$$f(x) = \begin{cases} 2 - 2x & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for alle andre } x. \end{cases}$$

- (a) Gi en enkel skisse av funksjonen  $f(x)$ . Forklar hvorfor denne  $f(x)$  er en sannsynlighetstetthet (probability density).
- (b) Anta at  $X$  er en stokastisk variabel (random variable) som har  $f(x)$  som sin sannsynlighetstetthet. Du skal nå finne tre sannsynligheter: for at  $X \leq 0.2$ ; for at  $X \leq 0.4$ ; og for at  $0.2 \leq X \leq 0.4$ .
- (c) Finn medianen i denne fordelingen.

### Oppgave 3

VI SKAL SE PÅ EN GANSKE ENKEL sannsynlighetsfordeling, nemlig den der verdiene som kan antas er 0.0, 0.2, 1.0, med sannsynligheter henholdsvis  $1/3$ ,  $1/3$ ,  $1/3$ .

- (a) Finn forventningsverdien  $\mu_X$  for denne fordelingen. Finn også standardavviket  $\sigma_X$  for fordelingen. (Svarene skal bli henholdsvis 0.4 og 0.432.)
- (b) Anta nå at  $X_1, X_2, X_3$  er tre uavhengige stokastiske variable, som hver har denne nettopp beskrevne fordelingen; spesielt kan hver av dem anta verdiene 0, 0.2, 1. Finn forventningsverdi og standardavvik for variablene  $Y = (X_1 + X_2)/2$  og  $Z = (X_1 + X_2 + X_3)/3$ .
- (c) Finn eksplisitt sannsynlighetsfordelingen for  $Y = (X_1 + X_2)/2$ .
- (d) Du har altså funnet i punkt (a) at forventningsverdien for  $X_i$ -ene er  $\mu_X = 0.4$ . Prøv å gi dette en tolkning, ved noen setninger som skal forklare for en person, med mindre dype kunnskaper enn det du selv har i statistikk og sannsynlighetsteori, hva det betyr at  $\mu_X = 0.4$ .

### Oppgave 4

MEN ER DET LØGN ELLER SANNHET? Anta (for denne oppgavens vedkommende) at menn utgjør 47% av befolkningen og snakker sant 64% av tiden, mens kvinner snakker sant 76% av tiden.

- (a) Hva er sannsynligheten for at en vilkårlig valgt person skal besvare et spørsmål sannferdig?
- (b) Hvis du får vite at denne vilkårlig valgte personen snakket sant, hva er da sannsynligheten for at vedkommende er en kvinne?

### Oppgave 5

VIL KOKAIN GJØRE FOSTERET MINDRE? Det anses etablert at kvinner som røyker underveis i svangerskapet statistisk sett får barn med lavere fødselsvekt. Forskere ved the Medical University of South Carolina i 1989 samlet inn data om gangen i graviditeten og om barnets helse for flere tusen kvinner, deriblant flere som hadde testet positivt på bruk av kokain. Fødselsvekten til 134 barn født av kokainbrukende kvinner hadde et gjennomsnitt på 2.733 kg og standardavvik på 0.599 kg. For normalgruppen omfattet undersøkelsen 5974 kvinner som ikke testet positivt for kokainbruk, med et gjennomsnitt på 3.118 kg og standardavvik på 0.672 kg.

Formuler en naturlig nullhypotese her, og test om den bør aksepteres eller forkastes. Formuler en konklusjon. Gjør rede for eventuelle antagelser du har benyttet deg av.

## Oppgave 6

KANDIDATENE GJØRES FØRST OPPMERKSOMME PÅ at en besvarelse av denne oppgaven kan gjøres langt kortere enn oppgaveteksten! Men for meg som forfatter den forklarende tekst er det altså nødvendig med en del utfyllende opplysninger, for å klarlegge sakens nødvendige faktorer.

Juliet B. Schor er professor i Women's Studies ved Harvard-universitetet, og har publisert flere bøker som har nådd ut til almenheten, bl.a. *The Overworked American: The Unexpected Decline of Leisure* fra 1992. Denne oppgaven skal berøre hennes hovedtese og hovedbudskap fra 1998-boken *The Overspent American: Why We Want What We Don't Need*. Dette hovedbudskapet er kort forklart at flere og flere amerikanere, på alle samfunnsnivåer, mer og mer lar seg lokke til å kjøpe enda mer og enda mer, for å følge med naboenes og ens bekjentes materielle utvikling (altså «to keep up with the Joneses»). Professor Schor er ikke snauere enn at hun også lanserer sitt eget ti-punktsprogram («Stopping the upward creep of desire») for å trekke USA ut av krisen.

Meget av bakgrunnen for hele hennes bok og budskap (og dermed også for hennes plan for å redde USA fra kjøpeeskaleringsspøkelsenes grusomme grep) er et par statistiske undersøkelser. I en av disse registrerte man

$$Y = \text{annual household saving,}$$

sammen med en rekke kovariater, for  $n = 834$  amerikanere i et bestemt populasjonssegment. Jeg skal ikke her gi den presise definisjonen av  $Y$ , men den skal altså være husstandens årlige sparebeløp, i US dollar, med visse modifikasjoner som bl.a. tar inn over seg alder og gjeld.

Schors viktigste anliggende er knyttet til spesialkovariaten  $x_{13}$ , som er personens selvopplevde finansielle status sammenlignet med ens egen (og like selvopplevde) referansegruppe. Hver person blir, etter å ha pekt ut ens naturlige referansegruppe (den gruppe av mennesker man mest naturlig sammenligner seg med), stilt spørsmålet: «Hvordan er din egen finansielle status, sammenlignet med de fleste av din referansegruppe?» Skalaen for  $x_{13}$  er da

$$x_{13} = \begin{cases} 1 & \text{much worse,} \\ 2 & \text{worse,} \\ 3 & \text{same,} \\ 4 & \text{better,} \\ 5 & \text{much better,} \end{cases}$$

og der det for øvrig var flest forekomster av «much worse» og «worse».

Alt dette resulterte i en multippel regresjonsanalyse som (blant annet) ga følgende resultater. Tabellen neste side gir for hver av tretten veldefinerte kovariater den estimerte regresjonskoeffisient  $b_j$ , dennes standard error (estimert standardavvik)  $SE = SE(b_j)$ , samt de tilhørende  $t$ -brøker  $t_j = b_j/SE(b_j)$ . (Tabellen gir også  $b_0$ ,  $SE$  og  $t_0$  for konstantleddet i regresjonsligningen. Skalaen for  $b_0$  er US dollar.)

	$b$	SE	$t$ -ratio	covariate
0	19995.000	15263.360	1.31	constant term
1	0.112	0.025	4.48	household income
2	0.025	0.050	0.50	permanent household income
3	0.016	0.019	0.84	household net worth
4	-3763.000	2187.791	-1.72	sex (1 = male, 2 = female)
5	-9916.000	6655.034	-1.49	age
6	1140.000	942.149	1.21	age-squared
7	-204.000	1700.000	-0.12	race
8	-174.000	424.390	-0.41	occupation
9	-1448.000	1366.038	-1.06	educational level
10	-1232.000	367.761	-3.35	number of dependents
11	629.000	767.073	0.82	satisfaction with income
12	-208.000	72.474	-2.87	hours per week watching TV
13	2963.000	773.629	3.83	relative financial status

- (a) Hvilke av de fjorten regresjonskoeffisientene  $\beta_0, \beta_1, \dots, \beta_{13}$  vil du si er signifikant betydningsfulle, for å forklare variasjonen i  $Y$ ? Begrunn ditt svar.
- (b) Lag et tilnærmet 95%-grads konfidensintervall for regresjonskoeffisienten  $\beta_{13}$ , og kommenter resultatet.