UNIVERSITETET I OSLO
ØKONOMISK INSTITUTT

Utsatt eksamen i: ECON2130 Statistikk 1

Eksamensdag: 04.06.2013

Tid for eksamen: kl. 09:00 - 12:00

Oppgavesettet er på 4 sider

Tillatte hjelpemiddel:
- Alle trykte og skrevne hjelpemiddel samt lommekalkulator

Eksamen blir vurdert etter ECTS-skalaen. A-F, der A er beste karakter og E er dårligste ståkarakter. F er ikke bestått.

Oppgave 1

Tabell 1 viser antall tannlegebesøk i løpet av et år for et utvalg av personer. Tabellen viser for eksempel at 4244 personer ikke hadde noen besøk hos tannlegen mens 267 hadde mer enn 4 besøk. Utvalget besto av 10218 personer i alt. (Tallene stammer fra en undersøkelse fra Nederland fra 1988.)

Tabell 1

Antall besøk i ett år	**Antall personer**
0	4244
1	1719
2	3337
3	461
4	190
> 4	267
Sum	10218

A. Skisser et histogram for fordelingen av antall tannlegebesøk basert på tallene i tabell 1. For enkelthets skyld kan du anta at de 267 personene som hadde mer enn 4 besøk, fordelte seg jevnt på 5, 6 og 7 besøk, dvs. med like mange på hver.

B. La X betegne antall tannlegebesøk i løpet av et år for en tilfeldig valgt person. Anta at sannsynlighetsfordelingen for X er gitt ved $f(x)$ i tabell 2. (Merk at tallene i tabell 2 ikke nødvendigvis er de samme som i histogrammet i punkt A.)

Tabell 2

x	0	1	2	3	4	5	6	7
$f(x)$	0.40	0.15	0.35	0.05	0.02	0.01	0.01	0.01

Beregn sannsynlighetene

i) $P(X = 4)$
ii) $P(X \geq 4)$
iii) $P\left[(X < 2) \cup (X > 4)\right]$
iv) $P\left[(X = 4) \mid (X \geq 4)\right]$

C. Beregn forventningen og variansen til X.

D. Anta at antall tannlegebesøk i året, X, for en tilfeldig person har forventning 1.26 og standardavvik 1.35. Tenk deg at vi observerer X for et tilfeldig utvalg på 10 000 personer. Hva er sannsynligheten, tilnærmet, for at gjennomsnittet av disse observasjonene skal falle mellom 1.25 og 1.27?

E. La X og Y betegne antall tannlegebesøk i løpet av et år for to tilfeldig valgte personer. Anta at begge følger fordelingen i tabell 2. Siden populasjonen er stor kan vi anta at X og Y er stokastisk uavhengige. Finn sannsynlighetene

i) $P\left[(X<2)\cup(Y<2)\right]$

ii) $P(X=Y)$

Hint: Begivenheten $(X=Y)$ kan skje ved at både X og Y blir lik 0 *eller* ved at begge blir lik 1 *eller* ved at begge blir lik 2 *eller* … osv.

Oppgave 2

A. La $X_1, X_2, \ldots, X_n$ være uavhengige og Poisson-fordelte med parameter λ.

i) Vis at gjennomsnittet, $\hat{\lambda}=\bar{X}$, er en forventningsrett estimator for λ.

ii) Hva blir variansen til $\hat{\lambda}$ uttrykt ved λ?

iii) Vi går tilbake til situasjonen i oppgave **1A**. Anta at antall tannlegebesøk for en tilfeldig valgt person er Poisson-fordelt med parameter λ. Estimer λ ut fra tallene i tabell 1. Som i oppgave **1A**, kan du anta at de med mer enn 4 besøk fordeler seg jevnt på 5, 6 og 7 besøk.

B. La situasjonen fortsatt være som beskrevet i punkt **A iii)**.

i) Beregn et (tilnærmet) 90% konfidensintervall for λ basert på tallene i tabell 1.

[**Hint.** Det kan vises (du behøver ikke gjøre det her) at $\dfrac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}/n}}$ er tilnærmet standard normalfordelt ($\sim N(0,1)$) uansett hva λ er. Utnytt dette resultatet til å lage konfidensintervallet.]

ii) Vi ønsker å teste hypotesen $H_0 : \lambda = 1.2$ mot $H_1 : \lambda \neq 1.2$. Sett opp et testkriterium med vilkårlig signifikansnivå α. Beregn p-verdien (signifikanssannsynligheten) for testen din basert på tallene i tabell 1 og kommenter svaret.

[**Hint.** Velg for eksempel $Z = \dfrac{\hat{\lambda} - 1.2}{\sqrt{\hat{\lambda}/n}}$ som testobservator. Andre forslag er også mulig.]

C. Det kan vises (du trenger ikke å gjøre det her) at forutsetningen i punkt **A iii)** at observasjonene i tabell 1 er trukket fra en Poisson-fordeling, er urealistisk. Dette reduserer troverdigheten av resultatene i punkt **B**. Som ledd i et forsøk på å forbedre modellen foreslås følgende:

Betrakt nå *utelukkende* dem som har 4 eller flere tannlegebesøk i løpet av et år. La U være antall tannlegebesøk i et år for en tilfeldig valgt person fra denne gruppen. Anta at $U = 4 + Y$ der Y er Poisson-fordelt med parameter μ.

i) Er det, etter din mening, noe som taler for og/eller i mot en slik antakelse?

ii) Hva blir forventningen og standardavviket til U, uttrykt ved μ?

iii) Vis at $P(U = 4) = e^{-\mu}$ og $P(U > 4) = 1 - e^{-\mu}$.

D. **i)** Forklar (intuitivt) hvorfor $\hat{\mu} = 0.87766$ kan være et rimelig anslag på μ ved å sammenligne sannsynlighetene i punkt **C iii)** med tilsvarende relative frekvenser beregnet på grunnlag av tabell 1.

ii) I oppgave **1A** fordelte vi de 267 personene, som hadde mer enn 4 tannlegebesøk, likt på 5, 6 og 7 besøk. Gjør rede for hvordan modellen i punkt **C** kan gi grunnlag for en annen fordeling av de 267 personene på 5, 6, 7, 8, … besøk. Angi de beregnete (estimerte) antall personer blant de 267 som hadde henholdsvis 5, 6 og 7 besøk hos tannlegen, med utgangspunkt i estimatet i punkt **i)**.

[**Hint.** Et uttrykk, $h(\mu)$, som avhenger av den ukjente μ, kan du her estimere ved rett og slett å erstatte μ med $\hat{\mu}$ i uttrykket, $h(\hat{\mu})$. Dette trenger du ikke å begrunne.]