Matematiska vetenskaper
Chalmers tekniska högskola

Tentamen: 2016-08-26 kl $8^{30} - 12^{30}$
Matematisk statistik för K, TMA074

Lärare: David Bolin, telefon 772 53 75. **Jour**: Adam Malik, telefon, 772 53 25

Tillåtna hjälpmedel: Formelsamling och valfri miniräknare med tömda minnen.

Korrekt, väl motiverad lösning ger poängen som är indikerad i parentes vid vardera uppgift. Totalt kan man få 40 poäng och betygsgränserna för betyg 3, 4 och 5 är 16, 24 och 32 poäng.

---

1. En taxi var inblandad i en smitningsolycka en natt i en stad där det finns två taxibolag, ett med gröna bilar och ett med blå. En polis försöker fastställa vilket bolag som var ansvarigt för olyckan. Ett vittne har pekat ut taxibilen som blå. Polisen undersöker vittnets tillförlitlighet under samma omständigheter som gällde under olyckan och kommer fram till att det identifierar rätt färg i 80% av fallen. Polisen undersöker också hur många bilar de två bolagen har och finner att 75% av alla taxibilar i staden är gröna.

   (a) Vad är sannolikheten för att bilen som var inblandad i olyckan tillhörde bolaget med blå taxibilar? (2p)

   (b) För att ett taxibolag ska kunna fällas krävs det att sannolikheten är hög för att bilen tillhörde det bolaget. Hur stor tillförlitlighet måste vittnet ha (hur ofta måste det identifiera rätt färg) för att sannolikheten för att taxin var blå ska vara minst 0.9? (3p)

2. (a) X_1 och X_2 är oberoende poissonfördelade variabler, $X_1 \sim$Po(1) och $X_2 \sim$Po(2), beräkna $\mathsf{P}(X_1 + X_2 > 2)$. (3p)

   (b) Antag att $X_i \sim$Po(1) är oberoende för $i = 1, \ldots, n$. Hur många X_i måste vi summera om vi vill att sannolikheten att $\sum_{i=1}^n X_i$ ska vara större än 2 är åtminstone 0.9? Alltså, hitta det minsta n så att $\mathsf{P}(\sum_{i=1}^n X_i > 2) > 0.9$. (2p)

3. För att motivera bygget av en ny vägbro sades det att minst 1000 personer per dygn skulle utnyttja bron. Efter bygget var klart ville man undersöka detta påstående och man räknade att 570 bilar passerade bron på ett dygn. Sedan tidigare vet man att antalet personer i en personbil i området har en fördelning som beskrivs av sannolikhetsfunktionen $p(k)$, med

k	1	2	3	4	5
$p(k)$	0.5	0.35	0.09	0.05	0.01

   och $p(k) = 0$ för alla andra värden på k. Beräkna sannolikheten att minst 1000 personer passerade över bron. (5p)

4. För att undersöka om kroppsvikt påverkar promillehalten i blodet lät man åtta kvinnor dricka varsitt glas vin varefter deras promillehalt i blodet mättes. Följande resultat erhölls:

Person (i)	1	2	3	4	5	6	7	8
Vikt (kg)	49	67	55	70	59	68	52	60
Halt (Promille)	0.210	0.125	0.182	0.109	0.165	0.122	0.196	0.161

   Från mätningarna kan man beräkna $\sum_i x_i = 480, \sum_i y_i = 1.27, \sum_i x_i^2 = 29224, \sum_i y_i^2 = 0.2112$ samt $\sum_i x_i y_i = 74.188$. Antag att promillehalten y_i hos en person med vikt x_i kan modelleras med hjälp av en regressionsmodell $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, där $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$.

   (a) Testa ifall vikten påverkar promillehalten. Kom ihåg att tydligt ange hypoteser och nivå för testet. (2p)

(b) Anna väger 60kg och hennes promillehalt kan därför skattas med $\mu_Y(60) = \beta_0 + 60\beta_1$. Om vi vill få en uppfattning om osäkerheten i skattningen kan vi antingen beräkna ett prediktionsintervall för $\mu_Y(60)$ eller ett konfidensintervall för Y. Förklara vad skillnaden mellan dessa två intervall är. (2p)

(c) Beräkna ett 95% prediktionsintervall för Annas promillehalt. (1p)

5. En butiksägare vill undersöka om placeringen av en viss vara i butiken påverkar hur väl den säljer. Antag att kunder i vanliga fall köper $X_i \sim \mathsf{N}(3, 1)$ antal varor, och att kunder istället köper $Y_i \sim \mathsf{N}(3 + \theta, 1)$ antal varor efter att butiksägaren bytt varans placering. För att testa påverkan av placeringen räknar företagaren hur många varor n kunder köper när den vanliga placeringen används, sedan mäter han också n varor med den nya placeringen. Han vill nu utföra hypotestestet

$$H_0 : \theta \leq 0$$
$$H_1 : \theta > 0$$

(a) Vid en mätning med $n = 5$ fås $\bar{x} = 4.36$ samt $\bar{y} = 5.78$. Utför hypotestestet på nivå 0.95, kan H_0 förkastas? (2p)

(b) Tre viktiga begrepp vid hypotestest är “typ 1 fel”, “typ 2 fel” samt “styrka”. Förklara dessa tre begrepp kortfattat. (3p)

(c) Butiksägaren tycker endast det är värt att göra omplaceringen permanent om θ är åtminstone 1. Givet att $\theta = 1$ och att testet utförs på nivå 0.95, hur stort måste n vara för att testets styrka ska vara åtminstone 0.9? (5p)

6. En bärplockare vill skatta antalet hjortron på en myr. Hon har noterat att en Poissonprocess kan användas för att modellera hur hjortron växer på myren, och hon behöver skatta intensiteten processen. Hon väljer ut två olika delar av myren, en med en area på $1m^2$ och en med en area på $3m^2$. Om hon räknar antalet hjortron i vardera område får han två oberoende stokastiska variabler, $X \sim$Po(θ) och $Y \sim$Po(3θ), där θ är den sökta intensiteten. Hon har tänkt ut två olika skattningsmetoder,

$$\theta_1^* = \frac{1}{2}\left(X + \frac{Y}{3}\right), \quad \theta_2^* = \frac{X+Y}{4}$$

och behöver nu hjälp med att avgöra vilken som är bäst.

(a) Visa att båda metoderna är väntevärdesriktiga. (3p)

(b) Avgör vilken som är mest effektiv, dvs vilken som har minst varians. (3p)

(c) Bärplockaren räknar antalet hjortron i de två områdena och finner $x = 9$ och $y = 18$. Skatta θ med den mest effektiva metoden och konstruera ett konfidensintervall med approximativ konfidensgrad 95%. (4p)

Lycka till!