Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1910/SF1925 TILLÄMPAD STATISTIK,
TORSDAG 13 APRIL 2023 KL 8.00–13.00.

Examinator: Björn-Olof Skytt, 08-790 8649.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik (utdelas vid tentamen), miniräknare.

Tentamen består av två delar, benämnda del I och del II. Del I består av uppgifterna 1-12. På denna del skall endast svar anges, antingen i form av ett numeriskt värde med tre värdesiffrors noggrannhet eller i form av val av ett av de möjliga svarsalternativen. Svaren anges på svarsblanketten. Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter (i svarsblanketten anges ordet Bonus). Studenter som är godkända på datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet Bonus). Detta gäller på ordinarietentan i januari 2023 och vid omtentamen i april 2023. Gränsen för godkänt är 9 poäng. Möjlighet att komplettera ges för tentander med 8 poäng.
Del II består av uppgifterna 13-16 och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på eller får komplettera del I och poäng på del II krävs för högre betyg än E. På denna del skall resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar skall förklaras och definieras och numeriska svar skall anges med minst två värdesiffrors noggrannhet. Studenter som är godkända på datorlaborationen får 3 bonuspoäng på del II på ordinarie tentamenstillfället och vid omtentamen i april 2023.

Tentamen kommer att vara rättad inom tre arbetsveckor(15 arbetsdagar) från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Del I

Uppgift 1

$P(A \cup B^*) = 0.85$. $P(A \cap B) = 0.25$. Bestäm $P(B)$.

A: 0.40

B: 0.45

C: 0.55

D: 0.60

Uppgift 2

$X \in U(-2, 2)$. $Y = X^2$. Bestäm standardavvikelsen $D(Y)$.

A: 1.19

B: 1.42

C: 1.79

D: 3.2

Uppgift 3

X och Y är oberoende stokastiska variabler. $X \in Bin(4, 0.5)$ och $Y \in Bin(4, 0.5)$.

Bestäm $P(min(X, Y) < 2)$

A: 0.098

B: 0.473

C: 0.527

D: 0.902

Uppgift 4

Åtta maskiner står uppställda i rad. De är i drift hela natten utan bevakning. Varje maskin har samma sannolikhet att stanna. När operatören kommer på morgonen har två av maskinerna stannat. Vad är sannolikheten att de två maskinerna som stannat står brevid varandra?

A: $\frac{1}{7}$

B: $\frac{2}{7}$

C: $\frac{1}{8}$

D: $\frac{1}{4}$

Uppgift 5

En nytt superbatteri för brandvarnare säges ha livstiden 10 år. Om vi med detta antar att det menas att väntevärdet av livstiden för batteriet är 10 år och att livstiden är exponentialfördelad, vad är då sannolikheten att vi måste byta batteri innan 10 år förflutit?

A: 0.368

B: 0.434

C: 0.500

D: 0.566

E: 0.632

Uppgift 6

Majas och Joels utgifter för kursmaterial (enhet: kr) under en månad kan ses som två oberoende stokastiska variabler. Majas utgifter/månad kan antas vara en normalfördelad s.v. med väntevärde 300 kr och standardavvikelse 50 kr och Joels utgifter/månad kan antas vara en normalfördelad s.v. med väntevärde 400 kr och standardavvikelse 70 kr. Vad är sannolikheten att Joels utgifter överstiger Majas under en slumpmässigt vald månad?

A: 0.123

B: 0.203

C: 0.797

D: 0.877

Uppgift 7

Antag att X är fördelat enligt $\text{Exp}(3\theta)$ medan Y är fördelat enligt $\text{Exp}(5\theta)$. Beräkna ML-skattningen för θ givet $x = 8.1$ och $y = 13.2$.

1. 0.0111
2. 0.0221
3. 0.0282
4. 0.0939

Uppgift 8

I en opinionsundersökning svarar 705 av 987 tillfrågade ja på frågan om de föredrar *Sourcream & Onion*-chips (från valfritt märke) över andra smaker. Resterande svarade nej. Om vi skattar andelen *Sourcream & Onion*-ätare p i befolkningen med $p^\star_{\text{obs}} = 705/987$, ange medelfelet för $p^\star_{\text{obs}}$.

1. 0.204
2. 0.0144
3. 0.452
4. 0.0318

Uppgift 9

Vi har följande observationer från ett normalfördelat stickprov:

$$1.67, 1.56, 1.76, 2.55, 0.78$$

Beräkna den övre gränsen för ett 95% tvåsidigt konfidensintervall.

1. 2.45
2. 2.22
3. 3.41
4. 2.26

Uppgift 10

Antag att vi har $X_1, X_2, \ldots, X_n$ från fördelningen $\mathrm{N}(\theta, 1.8)$ och vi vill pröva nollhypotesen $\theta = 0$ mot alternativet $\theta \neq 0$. Vårt test förkastar nollhypotesen om stickprovsmedelvärdet $\overline{x}$ skiljer sig tillräckligt från noll. Beräkna testets p-värde om vi observerar $\overline{x} = 0.9$ givet $n = 6$ observationer.

1. 0.889
2. 0.184
3. 0.222
4. 0.111

Uppgift 11

Efter rättningen av föregående tentamen (10 januari 2023) analyserades svaren för uppgift 5 (om binomialfördelningen) och uppgift 10 (om hypotestest) för att se om det fanns något samband. Följande data erhölls då.

Fel på båda:	17
Fel på uppgift 5, rätt på uppgift 10:	12
Rätt på uppgift 5, fel på uppgift 10:	57
Rätt på båda:	141

Nollhypotesen H_0 är att en student får rätt på uppgift 5 är oberoende av att studenten får rätt på uppgift 10. Vilket av följande stämmer?

1. H_0 kan ej förkastas med approximativ signifikansnivå 5% eller 1%.
2. H_0 kan förkastas med approximativ signifikansnivå 5% men ej 1%.
3. H_0 kan förkastas med approximativ signifikansnivå 1% men ej 5%.
4. H_0 kan förkastas med approximativ signifikansnivå 1% och 5%.

Uppgift 12

I en artikel 27 mars 2023 analyserar Dagens Nyheter överdödligheten i Sverige under coronapandemin 2020–2022. En aspekt som de lyfter är hur man beräknar den förväntade dödligheten för dessa år om pandemin inte skulle ha inträffat utifrån dödligheten i Sverige år 2015–2019.

År	2015	2016	2017	2018	2019
Antal döda per 10 000 invånare	92.7	91.3	90.9	89.7	84.8

Den första metoden tar helt enkelt medelvärdet över de fem senaste åren, men eftersom dödligheten följer en nedåtgående trend kan detta överskatta den förväntade dödligheten år 2020. Deras alternativ är att anpassa en linjär regressionsmodell och använda detta för att prediktera värdet år 2020. Med hur många döda per 10 000 invånare överskattar medelvärdet dödligheten jämfört med den linjära modellen?

1. 12.2
2. 5.22
3. 3.48
4. 8.70

Del II

Uppgift 13

Vid en statistisk kvalitetskontroll av partier innehållande 50 enheter används följande tvåstegsförfarande. Först dras 5 slumpmässiga enheter utan återläggning ur partiet. Om någon av dessa är defekt så avvisas partiet. Om ingen är defekt dras ytterligare 10 enheter slumpmässigt utan återläggning från de återstående 45. Om minst 2 av dessa är defekta så avvisas partiet.

(a) Vad är sannolikheten att avvisa ett parti som innehåller 7 defekta enheter? (5 p)

(b) Antag att partiet avvisades. Vad är då sannolikheten att det klarade det första steget men inte det andra? (5 p)

Uppgift 14

Låt X vara en stokastisk variabel med täthetsfunktion

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0, & \text{annars} \end{cases}$$

och låt Y vara likformigt fördelad över intervallet $[0, 1]$. Beräkna täthetsfunktionen för $Z = X + Y$ och rita upp den. (10 p)

Uppgift 15

På grund av den höga inflationen har den lokala pizzerian märkt av ett tapp i kunder de senaste månaderna. I ett försök att locka tillbaka dessa har pizzerian köpt annonser i en gratistidning som delas ut till alla hushåll i området. Veckorna innan annonsen publicerades sålde pizzerian följande antal pizzor per vecka

$$76, 76, 61, 76, 65$$

medan följande antal pizzor sålder under de veckorna som annonsen syntes i tidningen

$$72, 77, 84, 90, 66.$$

Antag att antalet sålda pizzor per vecka följer en poissonfördelning. Kan pizzeriaägaren rimligen dra slutsatsen att annonsen har haft önskad effekt, dvs. att väntevärdet för antal sålda pizzor per vecka har ökat? Motivera eventuella antaganden och approximationer. (10 p)

Uppgift 16

Vi har att $X \in Bin(5000, p)$ och $Y \in Bin(3000, 2p)$. Vidare har vi fått utfallen x och y. Anta att $H_0 : p = 0.02$ och $H_1 : p > 0.02$. Konstruera ett lämpligt hypotestest på risknivån 5% som tar både x och y i beaktande. Med andra ord, ange funktionen $f(x, y)$, så att $f(x, y) > 0$ om och endast om vi förkastar H_0.
Alla antaganden och eventuella approximationer skall motiveras. (10 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1910/SF1925 TILLÄMPAD STATISTIK, TORSDAG 13 APRIL 2023 KL 8.00–13.00.

Del I, Svar

1. A
2. A
3. C
4. D
5. E
6. D
7. B
8. B
9. A
10. C
11. D
12. B

Del I, Lösningsförslag:

Uppgift 1

Ritar man upp Venndiagrammet så ser man att $P(A \cup B^*) - P(A \cap B) = P(B^*) = 0.85 - 0.25 = 0.6$. Alltså är $P(B) = 0.4$.

Uppgift 2

$V(Y) = V(X^2) = E(X^4) - E^2(X^2)$
$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_{-2}^{2} x^2 \cdot \frac{1}{4}dx = \frac{16}{12} = \frac{4}{3}.$
$E(X^4) = \int_{-\infty}^{\infty} x^4 f_X(x)dx = \int_{-2}^{2} x^4 \cdot \frac{1}{4}dx = \frac{64}{20} = \frac{16}{5}.$
$V(Y) = V(X^2) = E(X^4) - E^2(X^2) = \frac{16}{5} - (\frac{4}{3})^2 = \frac{64}{45}$
$D(Y) = \sqrt{\frac{64}{45}} = 1.19$

Uppgift 3

$P(min(X,Y) < 2) = 1 - P(X \geq 2)P(Y \geq 2) = 1 - [1 - P(X \leq 1)][1 - P(X \leq 1)] = [se \quad tab \quad 6] = 1 - [1 - 0.3125]^2 = 0.527$

Uppgift 4

B - Den andra som stannar står brevid den första som stanar.
M- nr 1 är bland de 6 i mitten.
K- nr 1 är en av de 2 som står ytterst.
Här har vi lagen om total sannolikhet.

$P(B) = P(B|M)P(M) + P(B|K)P(K) = \frac{2}{7} \cdot \frac{6}{8} + \frac{1}{7} \cdot \frac{2}{8} = \frac{14}{56} = \frac{1}{4}$

Uppgift 5

Låt X vara batteriets livslängd. Enligt uppgiften är då X $\in$ Exp(λ) fördelad. Vi vet också att $E(X) = 10$ vilket ger att $\lambda = 1/10$. Att vi behöver byta batteriet innan 10 år har förflutit innebär att livslängden för batteriet är mindre än 10 så den sökta sannolikheten blir då

$$\mathrm{P}(\mathrm{X} < 10) = \int_0^{10} f_{\mathrm{X}}(x)dx = \int_0^{10} \frac{1}{10} e^{-\frac{1}{10}x} dx =$$
$$= \left[-e^{-\frac{1}{10}x}\right]_0^{10} = 1 - e^{-1} = 0.632$$

Svar: <u>Sannolikheten att behöva byta batteriet innan 10 år passerat är 0.632.</u>

Uppgift 6

Låt Majas utgifter vara M och Joels utgifter vara J. Sökt är $P(J - M > 0)$.
Vi har en linjärkombination av två oberoende Normalfördelade stokastiska variabler, så då är även $J - M$ Normalfördelad.

$E(J - M) = 400 - 300 = 100.$
$V(J - M) = V(J) + V(M) = 50^2 + 70^2 \Rightarrow D(J - M) = \sqrt{50^2 + 70^2} = 10\sqrt{74}.$

$$P(J - M > 0) = [\text{gör om till } N(0,1)] = P\left(\frac{J - M - 100}{10\sqrt{74}} > \frac{0 - 100}{10\sqrt{74}}\right) = 1 - \Phi\left(\frac{0 - 100}{10\sqrt{74}}\right) =$$

$$= 1 - \left[1 - \Phi\left(\frac{100}{10\sqrt{74}}\right)\right] = \Phi\left(\frac{100}{10\sqrt{74}}\right) = \Phi(1.16) = [\text{tab } 1] = 0.877$$

Uppgift 7

Vi har log-likelihoodfunktionen

$$\log L(\theta) = \log\left(3\theta e^{-3\theta x} \cdot 5\theta e^{-5\theta y}\right) = \log 15 + 2\log\theta - 3\theta x - 5\theta y$$

vars derivata med avseende på θ ger

$$\frac{2}{\theta} - 3x - 5y = 0$$

och vi får den kritiska punkten

$$\theta = \frac{2}{3x + 5y}.$$

Andraderivatan är lika med $-2/\theta^2$, så vi har ett global maximum. Stoppar vi in värdena får vi ML-skattningen 0.022.

Uppgift 8

Om X är antalet som svarar ja har vi att X är $\mathrm{Bin}(n,p)$-fördelad (där $n = 987$). Skattningen $p^\star = X/n$ har standardavvikelsen

$$D\left(p^\star\right) = \frac{1}{n} D\left(X\right) = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}.$$

Stoppar vi in skattningen $p^\star_{\mathrm{obs}} = 705/987$ får vi då medelfelet

$$\mathrm{d}(p^\star) \approx 0.0144.$$

Uppgift 9

Från datat har vi $\overline{x} = 1.664$ och $s^2 = 0.3966$, vilket ger

$$\overline{x} + \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1) = 1.664 + \frac{\sqrt{0.3966}}{\sqrt{5}} t_{0.025}(4) \approx 1.664 + 0.2816 \cdot 2.78 \approx 2.45.$$

Uppgift 10

Under nollhypotesen har vi att X är $\mathrm{N}(0, 1.8)$-fördelat och $\overline{X}$ har då fördelningen $\mathrm{N}(0, 1.8/\sqrt{6})$, dvs. $\mathrm{N}(0, 0.7348)$. Eftersom p-värdet motsvarar sannolikheten att (under nollhypotesen) erhålla ett mer extremt värde än det som observerats har vi

$$\mathrm{P}\left(|\overline{X}| > 0.9\right) = \mathrm{P}\left(\overline{X} < -0.9\right) + \mathrm{P}\left(\overline{X} > 0.9\right) = 1 - \mathrm{P}\left(\overline{X} < 0.9\right) + 1 - \mathrm{P}\left(\overline{X} < 0.9\right) = 2(1 - \mathrm{P}\left(\overline{X} < 0.9\right)).$$

Normalfördelningens räkneregler ger då att

$$\mathrm{P}\left(\overline{X} < 0.9\right) = \Phi\left(\frac{0.9}{0.7348}\right) \approx \Phi(1.22) \approx 0.8888.$$

Stoppar vi in detta i formeln för p-värdet får vi då $2(1 - 0.8888) = 0.2224$.

Uppgift 11

Ställer vi upp datan i tabell får vi

	Fel uppg. 5	Rätt uppg. 5	Radsumma
Fel uppg. 10	17	57	74
Rätt uppg. 10	12	141	153
Kolonnsumma	29	198	227

där cellen längst ner till höger är det totala antalet svar $N = 227$. Vi bildar nu kvadratsumman Q och får $Q = 10.25$. Då radsummorna uppfyller villkoret för χ^2-test har vi att Q ska vara approximativt $\chi^2(1)$-fördelad under nollhypotesen. De relevanta kvantilerna är $\chi^2_{0.05}(1) \approx 3.84$ samt $\chi^2_{0.01}(1) \approx 6.63$. Eftersom vårt värde på Q överstiger båda kan vi förkasta nollhypotesen (att svaren är oberoende) på båda signifikansgrader.

Uppgift 12

Om vi sätter år 2015 som $x = 0$, år 2016 som $x = 1$, osv. får vi

$$\sum_{i=1}^{5} x_i = 10 \quad \sum_{i=1}^{5} x_i^2 = 30$$

vilket ger

$$\overline{x} = 10/5 = 2 \quad s_{xx} = \sum_{i=1}^{5} (x_i - \overline{x})^2 = 30 - 5 \cdot 2^2 = 10.$$

På samma sätt ger datan oss att

$$\sum_{i=1}^{5} y_i = 449.4 \quad \sum_{i=1}^{5} x_i y_i = 881.4,$$

vilket ger

$$\overline{y} = 449.4/5 = 89.88 \quad s_{xy} = \sum_{i=1}^{5} (x_i - \overline{x})(y_i - \overline{y}) = 881.4 - 5 \cdot 2 \cdot 89.88 = -17.4.$$

Vi kan nu skatta lutningen till

$$\beta^\star_{\text{obs}} = \frac{s_{xy}}{s_{xx}} = -\frac{17.4}{10} = -1.74$$

medan konstanttermen fås till

$$\alpha^\star_{\text{obs}} = \overline{y} - \beta^\star_{\text{obs}}\overline{x} = 89.88 + 1.74 \cdot 2 = 93.36.$$

Vår prediktion för y i punkten $x = 5$ är då

$$93.36 - 1.74 \cdot 5 = 84.66.$$

Skillnaden mellan detta värde och medelvärdet är då $89.88 - 84.66 = 5.22$.

Del II, Lösningsförslag:

Uppgift 13

Vi inför händelserna A, $K1$, och $K2$, där A är att partiet avvisas, $K1$ är att partiet klarar första steget och $K2$ är att partiet klarar andra steget.

(a) Här söks $P(A)$. $P(A) = 1 - P(A^*) = 1 - P(K1 \cap K2) = [\text{ober}] = 1 - P(K1)P(K2) =$

$$= 1 - \left(\frac{\binom{7}{0}\binom{43}{5}}{\binom{50}{5}} \right) \left(\frac{\binom{7}{0}\binom{38}{10} + \binom{7}{1}\binom{38}{9}}{\binom{45}{10}} \right) = 1 - 0.454321395535 \cdot 0.505868493387 = 0.770173120127 \approx 0.77$$

(b) Antag att partiet avvisades. Vad är då sannolikheten att det klarade det första steget men inte det andra?

$$\text{Här söks } P((K1 \cap K2^*)|A) = \frac{P(K1 \cap K2^*)}{P(A)} = [\text{ober}] = \frac{P(K1)P(K2^*)}{P(A)} = \frac{P(K1)(1 - P(K2))}{P(A)} =$$

$$= \frac{\frac{\binom{7}{0}\binom{43}{5}}{\binom{50}{5}} \left(1 - \frac{\binom{7}{0}\binom{38}{10} + \binom{7}{1}\binom{38}{9}}{\binom{45}{10}} \right)}{1 - \left(\frac{\binom{7}{0}\binom{43}{5}}{\binom{50}{5}} \right) \left(\frac{\binom{7}{0}\binom{38}{10} + \binom{7}{1}\binom{38}{9}}{\binom{45}{10}} \right)} =$$

$$= \frac{0.454321395535 \cdot (1 - 0.505868493387)}{0.770173120127} = 0.291485783904 \approx 0.29$$

Alternativ lösning:

$$P((K1 \cap K2^*)|A) = \frac{P(K1 \cap K2^*)}{P(A)} = \frac{P(A) - P(K1^*)}{P(A)} = \left(1 - \frac{1 - P(K1)}{P(A)} \right) =$$

$$= [\text{med numeriska värden insatta enligt ovan}] = 0.291485783904 \approx 0.29$$

Uppgift 14

Vi har alltså

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0, & \text{annars} \end{cases}$$

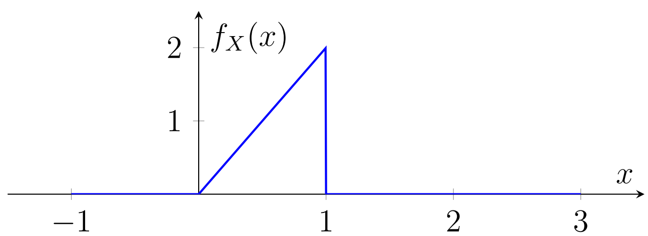
samt

$$f_Y(y) = \begin{cases} 1, & 0 \leq y \leq 1, \\ 0, & \text{annars} \end{cases}.$$

Eftersom täthetsfunktionen $f_Z(z)$ för $Z = X + Y$ ges av faltningen av $f_X(x)$ och $f_Y(y)$ har vi

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy \\ &= \int_0^1 f_X(z - y) dy \end{aligned}$$

eftersom $f_Y(y) = 1$ på intervallet $[0, 1]$.
Om vi ritar upp täthetsfunktionen $f_X(x)$



ser vi att det finns fyra sätt som intervallet $[z-1, z]$ kan korsa grafen: $z < 0$, $z \in [0, 1)$, $z \in [1, 2)$ och $z \geq 2$. Vi beräknar integralen för varje fall för sig.

- $z < 0$

$$f_Z(z) = \int_0^1 f_X(z-y)dy = \int_0^1 0dy = 0.$$

  eftersom $z - y < 0$ och $f_X(x) = 0$ för $x < 0$.

- $z \in [0, 1)$

$$f_Z(z) = \int_0^1 f_X(z-y)dy = \int_0^z f_X(z-y)dy = \int_0^z f_X(x)dx = \left[x^2\right]_0^z = z^2$$

  där vi utnyttjat att $f_X(z-y) = 0$ för $y \in [z, 1]$ och gjort variabelbytet $x = z - y$.

- $z \in [1, 2)$

$$f_Z(z) = \int_0^1 f_X(z-y)dy = \int_{z-1}^1 f_X(z-y)dy = \int_{z-1}^1 2xdx = \left[x^2\right]_1^{z-1} = 1-(z-1)^2 = 2z - z^2$$

  där vi utnyttjat att $f_X(z-y) = 0$ för $y \in [0, z]$ och gjort variabelbytet $x = z - y$.

- $z \geq 2$

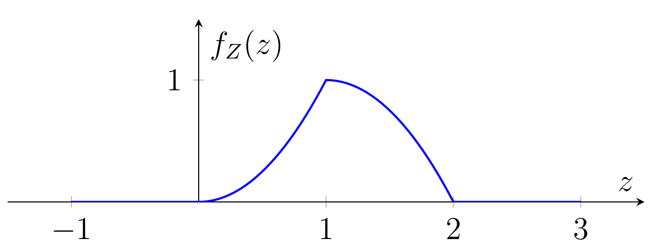
$$f_Z(z) = \int_0^1 f_X(z-y)dy = \int_0^1 0dy = 0$$

  eftersom $z - y > 1$ och $f_X(x) = 0$ för $x > 1$.

Vi har alltså

$$f_Z(z) = \begin{cases} z^2, & 0 \leq z < 1, \\ 2z - z^2, & 1 \leq z < 2, \\ 0, & \text{annars.} \end{cases}$$

Ritar vi upp täthetsfunktionen får vi



Uppgift 15

Låt θ_1 vara väntevärdet innan annonsen publicerades och θ_2 vara väntevärdet efter. Vi har då att de första fem veckorna följer fördelningen $\text{Po}(\theta_1)$ medan de resterande veckorna har fördelningen $\text{Po}(\theta_2)$. I båda fallen utgörs ML-skattningen av stickprovsmedelvärdet, vilket ger $\theta^{\star}_{1,\text{obs}} = \overline{x}_1 = 70.8$ samt $\theta^{\star}_{2,\text{obs}} = \overline{x}_2 = 77.8$. Eftersom dessa värden är tillräckligt stora (tumregeln i formelsamlingen anger att väntevärdet ska vara större än 15) kan vi använda normalapproximation, vilket ger att $\overline{X}_1$ approximativt har fördelningen $\text{N}(\theta_1, \sqrt{\theta_1/5})$ samt att $\overline{X}_2$ approximativt har fördelningen $\text{N}(\theta_2, \sqrt{\theta_2/5})$. Tillsammas ger detta att $\overline{X}_1 - \overline{X}_2$ har den approximativa fördelningen $\text{N}(\theta_1 - \theta_2, \sqrt{(\theta_1 + \theta_2)/5}$. Skattar vi standardavvikelsen till $\sqrt{(\theta^{\star}_{1,\text{obs}} + \theta^{\star}_{2,\text{obs}})/5} \approx 5.4516$ så får vi en övre gräns

$$(70.8 - 77.8) + \lambda_{0.05} \cdot 5.4516 \approx -7 + 8.9673 \approx 1.96$$

för ett 95% ensidigt konfidensintervall för väntevärdesskillnaden $\theta_1 - \theta_2$. Då detta intervall innehåller noll kan vi *inte förkasta* nollhypotesen $H_0 : \theta_1 = \theta_2$ till förmån för $H_1 : \theta_1 < \theta_2$.

Uppgift 16

$X \in Bin(n_x, p)$ och $Y \in Bin(n_y, 2p) = X \in Bin(5000, p)$ och $Y \in Bin(3000, 2p)$
En lämplig skattning av p fås m.h.a. MK-metoden.
$Q = (x - n_x p)^2 + (y - n_y 2p)^2$.

$$\frac{dQ}{dp} = 2(x - n_x p)(-n_x) + 2(y - 2n_y p)(-2n_y) = 0$$

$$n_x x - n_x^2 p + 2n_y y - 4n_y^2 p = 0$$

$$\Rightarrow p^*_{obs} = \frac{n_x x + 2n_y y}{n_x^2 + 4n_y^2} = \frac{5000x + 6000y}{61 \cdot 10^6}$$

$$E(p^*) = E\left(\frac{n_x X + 2n_y Y}{n_x^2 + 4n_y^2}\right) = \frac{n_x E(X) + 2n_y E(Y)}{n_x^2 + 4n_y^2} = \frac{n_x n_x p + 2n_y 2n_y p}{n_x^2 + 4n_y^2} = p$$

Således är skattningen väntevärdesriktig.
Om vi även antar att $n_x p(1-p) \geq 10$ och att $n_y 2p(1-2p) \geq 10$, så kan vi göra Normalapproximation. Vi antar alltså att p^* är approximativt N-fördelad,och då kan vi bilda ett konfidensintervall med approximativ konfidensgrad enligt §12.3.

Då behöver vi medelfelet.

$$V(p^*) = V\left(\frac{n_x X + 2n_y Y}{n_x^2 + 4n_y^2}\right) = [\text{vi antar ober}] = \frac{n_x^2 V(X) + 4n_y^2 V(Y)}{(n_x^2 + 4n_y^2)^2} =$$

$$= \frac{n_x^2 n_x p(1-p) + 4n_y^2 n_y 2p(1-2p)}{(n_x^2 + 4n_y^2)^2} = \frac{n_x^3 p(1-p) + 4n_y^3 2p(1-2p)}{(n_x^2 + 4n_y^2)^2} \Rightarrow$$

$$\Rightarrow [\text{medelfelet blir } D^*_{obs}(p^*) = \sqrt{\frac{n_x^3 p^*_{obs}(1-p^*_{obs}) + 4n_y^3 2p^*_{obs}(1-2p^*_{obs})}{(n_x^2 + 4n_y^2)^2}} =$$

$$= \frac{\sqrt{125 \cdot 10^9 p^*_{obs}(1-p^*_{obs}) + 216 \cdot 10^9 p^*_{obs}(1-2p^*_{obs})}}{61 \cdot 10^6}$$

Det nedåt begränsade konfidensintervallet som vi ska ha i detta fall blir $p^*_{obs} - D^*_{obs}(p^*)\lambda_\alpha$.

Vi förkastar H_0 om konfidensintervallet inte täcker över 0.02.
Dvs. om $p^*_{obs} - D^*_{obs}(p^*)\lambda_\alpha > 0.02$.
Alltså ska gälla att $p^*_{obs} - D^*_{obs}(p^*)\lambda_\alpha - 0.02 > 0$

$$\text{Dvs.} f(x,y) = p^*_{obs} - D^*_{obs}(p^*)\lambda_\alpha - 0.02 =$$

$$= \frac{n_x x + 2n_y y}{n_x^2 + 4n_y^2} - \sqrt{\frac{n_x^3 p^*_{obs}(1-p^*_{obs}) + 4n_y^3 2p^*_{obs}(1-2p^*_{obs})}{(n_x^2+4n_y^2)^2}} \cdot \lambda_{0.05} - 0.02 =$$

$$= \frac{5000x + 6000y}{61 \cdot 10^6} - \frac{\sqrt{125 \cdot 10^9 p^*_{obs}(1-p^*_{obs}) + 216 \cdot 10^9 p^*_{obs}(1-2p^*_{obs})}}{61 \cdot 10^6} \cdot 1.6449 - 0.02$$