Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1914/SF1915/SF1916 SANNOLIKHETSTEORI OCH STATISTIK, TISDAG 20 OKTOBER 2026 KL 8.00–13.00.

Kursledare SF1914/SF1915/SF1916: Björn-Olof Skytt,08 790 86 49

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 och varje korrekt svar ger 1 poäng. På denna del ska endast svar anges, i form av val av ett av de möjliga svarsalternativen på de uppgifter där svarsalternativ finns, annars skall svaret ges som ett numeriskt värde med *minst fyra* värdesiffrors noggrannhet. Svaren ska anges på svarsblanketten (utdelas vid tentamen). Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1–3, utan får tillgodoräkna sig dessa tre uppgifter (i svarsblanketten anges då ordet ”Bonus”). Studenter som är godkända på den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges då ordet ”Bonus”). Dessa tillgodoräknanden gäller för den här tentamen och vid omtentamen i december 2026. 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 ska resonemang och uträkningar vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar ska förklaras och definieras samt numeriska svar ska anges med minst tre värdesiffrors noggrannhet. Studenter som är godkända på den andra datorlaborationen får dessutom tre bonuspoäng på del II. Dessa bonuspoäng gäller för den här tentamen och vid omtentamen i december 2026.

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

Del I

Uppgift 1

Vi drar 5 kort slumpmässigt utan återläggning från en vanlig kortlek med 52 kort(där det finns 4 kungar). Vad är sannolikheten att vi får minst en kung?

A: ca 0.28

B: ca 0.34

C: ca 0.41

D: ca 0.72

Uppgift 2

I en viss stad är var nionde cykel en sportcykel och resten betecknas som vanliga cyklar. Sannolikheten att en sportcykel blir stulen under året är 4%, medan sannolikheten att en vanlig cykel blir stulen under året är 2%.
Vad är sannolikheten att en stulen cykel är en sportcykel?

Uppgift 3

De diskreta stokastiska variablerna Y och Z har den simultana sannolikhetsfunktionen $p_{Y,Z}(y,z) = yz/18, y = 1,2,3, z = 1,2$. Således antar Z alltså värdena 1 eller 2, medan Y antar värdena 1, 2, eller 3.
Bestäm väntevärdet $E(YZ)$.

A: 3.00

B: 3.47

C: 3.67

D: 3.89

Uppgift 4

Låt X och Y vara två oberoende stokastiska variabler sådana att $X \in Exp(1)$ och $Y \in Exp(1)$. Det gäller vidare att den stokastiska variabeln $Z = max(X,Y)$.
Beräkna $P(Z < 1)$.

A: 0.14

B: 0.40

C: 0.60

D: 0.86

Uppgift 5

Låt den stokastiska variabeln X vara likformigt fördelad mellan 0 och 1. Dvs. $X \in U(0,1)$. Det gäller vidare att den stokastiska variabeln $Y = X^2$.
Beräkna kovariansen C(X,Y).

Uppgift 6

Sannolikheten att en viss typ av elektronisk komponent skall ha en livslängd som är mindre än 1000 timmar är 0.25. Vad är den approximativa sannolikheten att av 20000 slumpvis valda sådana komponenter färre än 4900 kommer att ha en livslängd som är mindre än 1000 timmar?

A: ca 0.005

B: ca 0.025

C: ca 0.05

D: ca 0.10

Uppgift 7

X och Y är två oberoende stokastiska variabler. X är normalfördelad med väntevärdet 2μ och standardavvikelsen 2σ. Y är normalfördelad med väntevärdet μ och standardavvikelsen σ.

$$\mu^*_{obs} = \frac{x+2y}{4}$$

och

$$\hat{\mu}_{obs} = \frac{x+y}{3}$$

är två skattningar av μ. Vilket av de fyra följande alternativen stämmer?

A: Båda skattningarna är lika effektiva, dvs $V(\mu^*) = V(\hat{\mu})$.

B: Man kan inte avgöra vilken av skattningarna som är effektivast, eftersom minst en av skattningarna inte är väntevärdesriktig.

C: μ^*_{obs} är effektivare än $\hat{\mu}_{obs}$

D: $\hat{\mu}_{obs}$ är effektivare än μ^*_{obs}

Uppgift 8

En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \begin{cases} \frac{a}{x^{1+a}}, & \text{om } x \geq 1 \\ 0, & \text{annars} \end{cases}$$

Vi vet att a bara kan anta värdena 1, 2, 3 eller 4.
Bestäm Maximum-Likelihoodskattningen av a då vi fått observationerna $x_1 = 1.19$, $x_2 = 1.06$, och $x_3 = 2.16$.

A: 1

B: 2

C: 3

D: 4

Uppgift 9

Ett företag har under ett år utfärdat 5000 fakturor. Man vill se om det föreligger en systematisk skillnad mellan bokfört belopp och attesterat belopp. En revisor börjar med att slumpvis plocka ut 5 fakturor. Detta ger nedanstående tabell.

Faktura	1	2	3	4	5
Bokfört belopp (kkr)	300	720	520	200	127
Attesterat belopp (kkr)	300	520	520	200	157

Bilda ett tvåsidigt konfidensintervall med konfidensgraden 95% för skillnaden mellan väntevärdet för bokförda belopp och väntevärdet för attesterade belopp baserat på tabellens data och ange nedre gränsen. Utgå från en modell baserad på normalfördelningsantaganden.

A: -300.1

B: -82.5

C: -55.3

D: -48.1

Uppgift 10

Låt datamängden 23, 42, 55, 69, 39 vara given. Varje data anses vara ett utfall av en stokastisk variabel X_i, där X_i:na antas vara oberoende och Poissonfördelade med väntevärde μ. D.v.s. varje $X_i \in Po(\mu)$. Eftersom X_i:na är Poissonfördelade skattar vi $E(X_i) = \mu$ med $\bar{x}$ och $D(X_i) = \sqrt{\mu}$ med $\sqrt{\bar{x}}$. Ange nedre gränsen för det enkelsidigt nedåt begränsade konfidensintervallet I_μ för μ som har den approximativa konfidensgraden 95%.

A: ca 12.96

B: ca 32.36

C: ca 34.49

D: ca 40.63

Uppgift 11

Antag att den stokastiska variabeln X är För-första-gången-fördelad med sannolikheten p. Dvs $X \in fffg(p)$. Vi ställer upp $H_0 : p = 0.1$. Vi vill testa nollhypotesen H_0 mot alternativet $H_1 : p = 0.6$ och förkastar H_0 till förmån för mothypotesen H_1 om observationen $x < 4$.
Bestäm testets styrka.

A: 0.081

B: 0.271

C: 0.784

D: 0.936

Uppgift 12

Ett bibliotek försöker få bukt med sina problem med sent återlämnade böcker genom att höja avgiften som måste betalas vid sen återlämning. Efter att ha provat olika avgifter kom de fram till följande resultat:

Avgift (kr), x_i	0 kr	5 kr	10 kr	20 kr
Snittid för sen återlämning (dagar), y_i	11.2	9.7	10.1	8.1

Från datan har vi

$$\sum_i x_i = 35 \quad \sum_i x_i^2 = 525 \quad \sum_i y_i = 39.1 \quad \sum_i y_i^2 = 387.15 \quad \sum_i x_i y_i = 311.5,$$

vilket ger $\overline{x} = 8.75$, $\overline{y} = 9.775$ och

$$s_{xy} = 311.5 - 4 \cdot 8.75 \cdot 9.775 = -30.625$$

samt

$$s_{xx} = 525 - 4 \cdot 8.75^2 = 218.75$$

Sätt upp en enkel linjär regressionsmodell (med normalfördelade residualer) för återlämningstiden Y som funktion av avgiften x och bilda ett konfidensintervall I_β med konfidensgrad 95% för lutningen β. Vilket påstående nedan är korrekt?

A: Eftersom $0 \in I_\beta$ kan vi inte förkasta att återlämningstiden är oberoende av avgiften på risknivån 5%.

B: Eftersom $0 \in I_\beta$ kan vi förkasta att återlämningstiden är oberoende av avgiften på risknivån 5%.

C: Eftersom $0 \notin I_\beta$ kan vi förkasta att återlämningstiden är oberoende av avgiften på risknivån 5%.

D: Eftersom $0 \notin I_\beta$ kan vi inte förkasta att återlämningstiden är oberoende av avgiften på risknivån 5%.

Var god vänd!

Del II

Uppgift 13

Den tredimensionella kontinuerliga stokastiska variabeln (X, Y, Z) är likformigt fördelad i tetraedern

$$T = \{(x, y, z) \in \mathbb{R}^3 : x \geq 0,\ y \geq 0,\ z \geq 0,\ x + y + z \leq 3\}\,.$$

Beräkna $P(X + Y + Z \leq 2)$. (10 p)

Uppgift 14

Ett problem med statistiska undersökningar är ofta den låga svarsfrekvensen. Därför gjordes en undersökning om man genom att ge något incitament för att svara kunde förändra svarsfrekvensen på enkäter. Man sände därför ut 98 enkäter där man utlovade 10 trisslotter till den som svarade vilket ledde till att 66 personer svarade på enkäten. samtidigt sände man ut 110 likadana enkäter där det saknades erbjudande om trisslotter vilket ledde till 75 besvarade enkäter.

Ta fram ett tvåsidigt konfidensintervall med approximativ konfidensgrad 95% för skillnaden mellan andelen som svarar på enkäten om de får erbjudande om trisslotter och andelen som svarar på enkäten om de inte får erbjudande om trisslotter.
Svara sedan på om man kan förkasta att det inte är någon signifikant skillnad på risknivån 5% mellan andelen som svarar på enkäten om de får erbjudande om trisslotter och andelen som svarar på enkäten om de inte får erbjudande om trisslotter.
Ange dina hypoteser noga och motivera noga din slutsats. (10 p)

Var god vänd!

Uppgift 15

En ekolog har delat in ett område i 48 exakt lika stora kvadrater. Hon vill undersöka om antalet tallar som finns i varje kvadrat kan anses komma från en Poissonfördelning med väntevärdet 2 och har fått följande resultat.

Antal tallar	0	1	2	3	minst 4
Antal kvadrater	9	9	10	14	6

Kan man på signifikansnivån 5% förkasta att antalet tallar Y som finns i varje kvadrat kan anses komma från en Poissonfördelning med väntevärdet 2? Ange noga dina hypoteser. (10 p)

Uppgift 16

Små partiklar som svävar i atmosfären (kallade aerosoler) har radier som kan beskrivas med en lognormalfördelning. Dvs. om den stokastiska variabeln X är partikelns radie (i mikrometer) så gäller att den stokastiska variabeln $Y = lnX \in N(\mu, \sigma)$.

(a) Vad är sannolikheten att en atmosfärisk partikel har en radie som överstiger 0.12 mikrometer om $\mu = -2.62$ och $\sigma = 0.788$? (2 p)

(b) Härled täthetsfunktionen $f_X(x)$ som funktion av μ och σ. Uträkningarna skall visas. Man ska ej hänvisa till miniräknare. (3 p)

(c) Härled väntevärdet $E(X)$ för partikelns radie som funktion av μ och σ. Uträkningarna skall visas. Man ska ej hänvisa till miniräknare. (5 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1914/SF1915/SF1916
SANNOLIKHETSTEORI OCH STATISTIK,
TISDAG 20 OKTOBER 2026 KL 8.00–13.00.

Del I, Svar.

1. B

2. 0.2

3. D

4. B

5. $1/12 = 0.08333$

6. C

7. C

8. C

9. B

10. D

11. D

12. A

Del I, Lösningsförslag.

Uppgift 1

$$P(\text{minst en kung}) = 1 - P(\text{ingen kung}) = 1 - \frac{\binom{48}{5}\binom{4}{0}}{\binom{52}{5}} = 1 - \frac{1712304}{2598960} \approx 0.34$$

Uppgift 2

Vi inför händelserna $A \Leftrightarrow$ Sportcykel, $B \Leftrightarrow$ Vanlig cykel, $C \Leftrightarrow$ Stulen cykel.
Vi söker $P(A|C)$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A \cap C)}{P(A \cap C) + P(B \cap C)} = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)} =$$
$$= \frac{0.04 \cdot \frac{1}{9}}{0.04 \cdot \frac{1}{9} + 0.02 \cdot \frac{8}{9}} = \frac{0.04}{0.20} = 0.2$$

Uppgift 3

$$E(YZ) = \sum_{\text{alla y,z}} yz p_{Y,Z}(y,z) = \sum_{\text{alla y,z}} \frac{(yz)^2}{18} = \frac{1}{18}[(1\cdot1)^2+(2\cdot1)^2+(3\cdot1)^2+(1\cdot2)^2+(2\cdot2)^2+(3\cdot2)^2] = 3.89$$

Uppgift 4

$P(Z < 1) = P(max(X,Y) < 1) = P(X < 1 \cap Y < 1) = [\text{ober}] = P(X < 1)P(Y < 1)$

$$P(X < 1) = P(Y < 1) = \int_0^1 \lambda e^{-\lambda t} dt = \int_0^1 e^{-t} dt = 1 - e^{-1} \Rightarrow P(Z < 1) = (1 - e^{-1})^2 = 0.40$$

Uppgift 5

$C(X,Y) = E(XY) - E(X)E(Y) = [\text{i detta fall}] = E(X^3) - E(X)E(X^2)$

$$f_X(x) = \begin{cases} \frac{1}{1-0}, & \text{om } 0 \le x \le 1 \\ 0, & \text{annars} \end{cases}$$

$$E(X^3) = \int_0^1 x^3 \frac{1}{1-0} dx = \frac{1}{4}$$

$$E(X^2) = \int_0^1 x^2 \frac{1}{1-0} dx = \frac{1}{3}$$

$$E(X) = \int_0^1 x \frac{1}{1-0} dx = \frac{1}{2}$$

$$\Rightarrow C(X,Y) = \frac{1}{4} - \frac{1}{3} \cdot \frac{1}{2} = \frac{2}{24} = 0.08333$$

Uppgift 6

X är antal dragna komponenter som har en livslängd som är mindre än 1000 timmar.
Här gäller att $X \in Bin(n,p) = Bin(20000, 0.25)$.
$Bin(n,p)$ approximeras av $N(np, \sqrt{np(1-p)})$ om $np(1-p) \geq 10$.

Här är $np(1-p) = 5000 \cdot 0.75 = 3750$ så $X \approx N(5000, \sqrt{3750})$

$$P(X < 4900) = P\left(\frac{X-5000}{\sqrt{3750}} < \frac{4900-5000}{\sqrt{3750}}\right) = \Phi\left(\frac{4900-5000}{\sqrt{3750}}\right) \approx 1-\Phi(1.63) \approx 1-0.9489 = 0.05$$

Uppgift 7

X och Y är två oberoende stokastiska variabler. X är normalfördelad med väntevärdet 2μ och standardavvikelsen 2σ. Y är normalfördelad med väntevärdet μ och standardavvikelsen σ.

$$\mu^*_{obs} = \frac{x+2y}{4}$$

och

$$\hat{\mu}_{obs} = \frac{x+y}{3}$$

är två skattningar av μ.

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

$$E(\hat{\mu}) = E\left(\frac{X+Y}{3}\right) = \frac{2\mu+\mu}{3} = \mu$$

Båda skattningarna är alltså väntevärdesriktiga.

$$V(\mu^*) = V\left(\frac{X+2Y}{4}\right) = \frac{4\sigma^2+4\sigma^2}{16} = \frac{\sigma^2}{2}$$

$$V(\hat{\mu}) = V\left(\frac{X+Y}{3}\right) = \frac{4\sigma^2+\sigma^2}{9} = \frac{5\sigma^2}{9}$$

Således gäller att $V(\mu^*) < V(\hat{\mu})$. Alltså gäller att μ^*_{obs} är effektivare än $\hat{\mu}_{obs}$. Svarsalternativ C är det rätta.

Uppgift 8

Se §9.1 Det a som maximerar$L(a)$ väljs till Maximum-Likelihood-skattningen av a.

$$L(a) = \frac{a}{x_1^{1+a}} \cdot \frac{a}{x_2^{1+a}} \cdot \frac{a}{x_3^{1+a}} = \frac{a^3}{(x_1 \cdot x_2 \cdot x_3)^{1+a}} = \frac{a^3}{(1.19 \cdot 1.06 \cdot 2.16)^{1+a}}$$

Eftersom a endast kan anta värdena 1,2,3,och 4 ser vi vad $L(a)$ blir för dessa värden på a.
$L(1) = 0.1347, L(2) = 0.3955, L(3) = 0.4899, L(4) = 0.4262$
Dvs. $a = 3$ maximerar $L(a)$ och alltså är 3 Maximum-Likelihood-skattningen av a.

Uppgift 9

Låt $x_1, \ldots, x_5$ och $y_1, \ldots, y_5$ beteckna bokförda respektive de attesterade beloppen. Vi har här *stickprov i par*. Med normalfördelningsantaganden blir alltså modellen att med $z_i = x_i - y_i$ är $z_1, \ldots, z_5$ oberoende observationer från $N(\mu, \sigma)$. Uppgiften är att undersöka om en systematisk skillnad finns, dvs om $\mu \neq 0$.
De $n = 5$ z-värdena är 0, 200, 0, 0,-30, vilket ger $\sum z_i = 170$, $\overline{z} = 34$,

$$s = \sqrt{\frac{\sum (z_i - \bar{z})^2}{n-1}} = \sqrt{\frac{34^2 + 34^2 + 166^2 + 34^2 + (-64)^2}{4}} = \sqrt{8788}$$

Nu bildar vi ett tvåsidigt konfidensintervall för μ
och får enl. §12.2 $I_\mu = \bar{z} \pm \frac{s}{\sqrt{n}} t_{0.025}(4) = 34 \pm \frac{\sqrt{8788}}{\sqrt{5}} \cdot 2.78 = 34 \pm 116.5$
Alltså blir den nedre gränsen - 82.5.

Uppgift 10

Eftersom väntevärdet $E(X) = \mu$ så skattar vi μ med $\bar{x} = \frac{23+42+55+69+39}{5} = \frac{228}{5} 045.6$

Eftersom summan av oberoende Poissonfördelade variabler
är Poissonfördelad gäller i vårt fall att $\sum X_i \in Po(5\mu)$.

Eftersom $5\mu^*_{obs} = 228 \geq 15$ så är $Y = \sum X_i$ approximativt Normalfördelad och då är även $\bar{X}$ det eftersom $\bar{X}$ då är en linjärkombination av en Normalfördelad stokastisk variabel. Alltså kan vi bilda ett konfidensintervall med approximativ konfidensgrad enligt § 12.3.

Då kan vi approximera $Po(\mu)$ till $N(\mu, \sqrt{\mu})$ enl §6 i Formelsamlingen och därmed bilda ett konfidensintervall med approximativ konfidensgrad enl. §12.3.

Enligt §12.3 blir då det dubbelsidiga konfidensintervallet $I_\mu = \mu^*_{obs} \pm D^*_{obs}(\mu^*) \cdot \lambda_{\frac{\alpha}{2}}$.

$$\mu^*_{obs} = \bar{x} \Rightarrow D(\mu^*) = D(\bar{X}) = \frac{D(X)}{\sqrt{5}} = \frac{\sqrt{\mu}}{\sqrt{5}} \Rightarrow D^*_{obs}(\mu^*) = \frac{\sqrt{\bar{x}}}{\sqrt{5}}$$

I det enkelsidgt nedåt begränsade konfidensintervallet byter vi $\alpha/2$ mot α i den nedre gränsen och sätter den övre gränsen till ∞.
Dvs. det enkelsidiga konfidensintervallet blir

$$I_\mu = (\bar{x} - \frac{\sqrt{\bar{x}}}{\sqrt{5}} \cdot \lambda_{0.05}, \infty) = (46.5 - \frac{\sqrt{46.5}}{\sqrt{5}} \cdot 1.6449, \infty) \Rightarrow \text{nedre gränsen blir } 40.63$$

Uppgift 11

Styrkan hos testet är $P(\text{förkasta } H_0)$ om H_1 är sann.

Dvs. vi ska räkna ut $P(X < 4)$ då $p = 0.6$

vilket är $P(X = 1) + P(X = 2) + P(X = 3) = 0.6 + 0.6 \cdot 0.4 + 0.6 \cdot 0.4^2 = 0.936$

Uppgift 12

Se §13

$$I_\beta = [\text{§13.2 och §13.3}] = \beta^*_{obs} \pm t_{p/2}(n-2)\frac{s}{\sqrt{s_{xx}}}$$

$$\beta^*_{obs} = [\text{§13.1}] = \frac{s_{xy}}{s_{xx}} = \frac{-30.625}{218.75} = -0.14$$

$$t_{p/2}(n-2) = t_{0.05/2}(4-2) = t_{0.025}(2) = 4.30$$

$$\text{se §13.3 } (n-2)s^2 = s_{yy} - \frac{s^2_{xy}}{s_{xx}}$$

$$s_{yy} = \sum y_i^2 - n(\bar{y})^2 = 387.15 - 4 \cdot 9.775^2 = 4.9475$$

Detta ger att

$$2s^2 = 4.9475 - \frac{30.625^2}{218.75} = 0.66 \Rightarrow s = \sqrt{0.33}$$

Allt detta ger att

$$I_\beta = -0.14 \pm 4.30\frac{\sqrt{0.33}}{\sqrt{218.75}}$$

$$I_\beta = -0.14 \pm 0.167$$

$0 \in I_\beta$ Alltså kan vi inte förkasta att återlämningstiden är oberoende av avgiften på risknivån 5%. A är rätt svar.

Del II, Lösningsförslag.

Uppgift 13

För en tetraeder $T_a = (x, y, z) \in \mathbb{R}^3 : x \geq 0,\ y \geq 0,\ z \geq 0,\ x + y + z \leq a\ a > 0$, ges volymen av:

$$\mathrm{Vol}(T_a) = \int_0^a dx \int_0^{a-x} dy \int_0^{a-(x+y)} dz = \int_0^a dx \int_0^{a-x} (a - x - y)\, dy = \int_0^a \frac{(a-x)^2}{2}\, dx = \frac{a^3}{6}$$

Alltså är

$$\mathrm{Vol}(T_3) = \frac{3^3}{6} = \frac{27}{6}, \qquad \mathrm{Vol}(T_2) = \frac{2^3}{6} = \frac{8}{6}$$

Eftersom (X, Y, Z) är likformigt fördelad över T_3, så är tätheten $f_{X,Y,Z}(x, y, z)$ konstant, och sannolikheten för en delregion ges därför av volymkvoten:

$$P(X + Y + Z \leq 2) = \frac{\mathrm{Vol}(T_2)}{\mathrm{Vol}(T_3)} = \frac{2^3/6}{3^3/6} = \left(\frac{2}{3}\right)^3 \simeq 0.2963$$

Uppgift 14

Vår modell är att antalet som svarar på enkäten om de får trisslotter är en stokastisk variabel X där $X \in Bin(98, p_x)$ och att antalet som svarar på enkäten om de inte får trisslotter är en stokastisk variabel Y där $Y \in Bin(110, p_y)$.
Våra hypoteser är:$H_0 : p_x = p_y$,$H_1 : p_x \neq p_y$
Vi vill beräkna ett konfidensintervall med den approximativa konfidensgraden 95% för $p_x - p_y$. Vi skattar p_x med $p^*_{x_{obs}} = \frac{x}{n_x} = \frac{66}{98}$ och p_y med $p^*_{y_{obs}} = \frac{y}{n_y} = \frac{75}{110}$.
Både $n_x p^*_{x_{obs}}(1 - p^*_{x_{obs}}) = 98\frac{66}{98}(1 - \frac{66}{98}) = 21.55 \geq 10$ och
$n_y p^*_{y_{obs}}(1 - p^*_{y_{obs}}) = 110\frac{75}{110}(1 - \frac{75}{110}) = 23.86 \geq 10$. I båda fallen kan vi alltså approximera binomialfördelningarna med normalfördelningar.
Vi får att X är approximativt $N(n_x p_x, \sqrt{n_x p_x (1 - p_x)})$ och
Y är approximativt $N(n_y p_y, \sqrt{n_y p_y (1 - p_y)})$ vilket ger att
p^*_x är approximativt $N(p_x, \sqrt{p_x(1 - p_x)/n_x})$ och p^*_y är approximativt $N(p_y, \sqrt{p_y(1 - p_y)/n_y})$ vilket ger att $p^*_x - p^*_y$ är approximativt

$$N\left(p_x - p_y, \sqrt{\frac{p_x(1 - p_x)}{n_x} + \frac{p_y(1 - p_y)}{n_y}}\right)$$

som med approximativa metoden (FS 12.3) ger ett approximativt 95%-igt konfidensintervall för $p_x - p_y$

$$I_{p_x - p_y} = p^*_{x_{obs}} - p^*_{y_{obs}} \pm \lambda_{0.025} \sqrt{\frac{p^*_{x_{obs}}(1 - p^*_{x_{obs}})}{n_x} + \frac{p^*_{y_{obs}}(1 - p^*_{y_{obs}})}{n_y}}$$

$$I_{p_x - p_y} = \frac{66}{98} - \frac{75}{110} \pm 1.96 \sqrt{\frac{66 \cdot 32}{98^3} + \frac{75 \cdot 35}{110^3}} \approx -0.008 \pm 0.127.$$

Eftersom 0 ingår i konfidensintervallet kan vi inte fökasta $H_0 : p_x = p_y$ på risknivån 5%.

Uppgift 15

Här har vi test av given fördelning där nollhypotesen H_0 är att antalet tallar per kvadrat = $Y \in Po(2)$.
Se §14.3 Vi har här 5 resultat: $A_1 : 0$ tallar, $A_2 : 1$ tall, $A_3 : 2$ tallar, $A_4 : 3$ tallar, och A_5:minst 4 tallar.
$x_1 = 9, x_2 = 9, x_3 = 10, x_4 = 14, x_5 = 6$
$p_1 = P(Y = 0) = \frac{\mu^0}{0!} \cdot e^{-\mu}$
$p_2 = P(Y = 1) = \frac{\mu^1}{1!} \cdot e^{-\mu}$
$p_3 = P(Y = 2) = \frac{\mu^2}{2!} \cdot e^{-\mu}$
$p_4 = P(Y = 3) = \frac{\mu^3}{3!} \cdot e^{-\mu}$
$p_5 = 1 - p_1 - p_2 - p_3 - p_4$

Detta ger att

$$np_1 = 48 \cdot \frac{2^0}{0!} \cdot e^{-2} = 6.496$$

$$np_2 = 48 \cdot \frac{2^1}{1!} \cdot e^{-2} = 12.992$$

$$np_3 = 48 \cdot \frac{2^2}{2!} \cdot e^{-2} = 12.992$$

$$np_4 = 48 \cdot \frac{2^3}{3!} \cdot e^{-2} = 8.661$$

$$np_5 = 48(1 - p_1 - p_2 - p_3 - p_4) = 6.858$$

Alla $np_i \geq 5$.

$$Q = \frac{(9 - 6.496)^2}{6.496} + \frac{(9 - 12.992)^2}{12.992} + \frac{(10 - 12.992)^2}{12.992} + \frac{(14 - 8.661)^2}{8.661} + \frac{(6 - 6.858)^2}{6.858} = 6.279$$

Eftersom $Q = 6.279 < \chi^2_{0.05}(4) = 9.49$ så kan vi inte förkasta nollhypotesen H_0 på risknivån 5% som är att antalet tallar per kvadrat är Poissonfördelat med väntevärde 2.

Uppgift 16

a)

$$P(X > 0.12) = P(lnX > ln0.12) = P(\frac{y + 2.62}{0.788} > \frac{ln0.12 + 2.62}{0.788}) =$$

$$= 1 - \Phi(\frac{ln0.12 + 2.62}{0.788}) \approx 1 - \Phi(0.63) \approx 1 - \Phi(0.64) \approx 0.26$$

b)

$$F_X(x) = P(X \leq x) = P(e^Y \leq x) = P(Y \leq lnx) =$$

$$= P(\frac{Y - \mu}{\sigma} \leq \frac{lnx - \mu}{\sigma}) = \Phi(\frac{lnx - \mu}{\sigma})$$

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}\Phi(\frac{lnx - \mu}{\sigma}) = \frac{1}{x\sigma}\phi(\frac{lnx - \mu}{\sigma}) = \frac{1}{x\sigma} \cdot \frac{1}{\sqrt{2\pi}}e^{\frac{-(lnx-\mu)^2}{2\sigma^2}}; x > 0$$

c)

$$E(X) = E(e^Y) = \int_{-\infty}^{\infty} e^y \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy =$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(y-\mu)^2 + y \cdot 2\sigma^2}{2\sigma^2}} dy = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-y^2 - \mu^2 + 2y\mu + y \cdot 2\sigma^2}{2\sigma^2}} dy =$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(y-(\mu+\sigma^2))^2 + 2\mu\sigma^2 + \sigma^4}{2\sigma^2}} dy = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(y-(\mu+\sigma^2))^2}{2\sigma^2}} \cdot e^{\frac{2\mu\sigma^2 + \sigma^4}{2\sigma^2}} dy =$$

$$= e^{\frac{2\mu\sigma^2 + \sigma^4}{2\sigma^2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(y-(\mu+\sigma^2))^2}{2\sigma^2}} dy = e^{\mu + \frac{\sigma^2}{2}} \cdot 1 = e^{\mu + \frac{\sigma^2}{2}}$$