



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1920/SF1921 SANNOLIKHETSTEORI OCH STATISTIK,
FREDAG 6 MARS 2026 KL 8.00–13.00.

Examinator SF1920/SF1921: Mykola Shykula, 076-830 08 01.

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 ska endast svar anges. Svaren ska anges på svarsblanketten (utdelas vid tentamen). Varje uppgift kan ge antingen 1 eller 0 poäng (korrekt svar ger 1 poäng, felaktigt svar ger 0 poäng). Ange ett (och endast ett) svarsalternativ (A–D) då dessa finns. Annars, svara med minst **fyra värdesiffrors** noggrannhet. 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 Laboration 2 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 juni 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 och förstå. Införda beteckningar ska förklaras och definieras samt slutgiltiga numeriska svar ska anges med minst **fyra värdesiffrors** noggrannhet. Studenter som är godkända på Laboration 2 får dessutom tre bonuspoäng på del II. Dessa bonuspoäng gäller för den här tentamen och vid omtentamen i juni 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

För händelserna A och B gäller att $P(A \cap B^*) = 0.2$, $P(A^* \cap B) = 0.4$ och $P(A \cup B) = 0.9$. Bestäm $P(B | A)$.

A: 0.40

B: 0.43

C: 0.57

D: 0.60

Uppgift 2

På julbordet ligger tre skivor kallrökt lax, fyra skivor gravad lax och fem skivor varmrökt lax. Lille Nisse tar tre skivor helt på måfå. Vad är sannolikheten att Lille Nisse får tre skivor gravad lax?

Uppgift 3

Låt X_1 , X_2 och X_3 vara tre oberoende stokastiska variabler sådana att $X_1 \in Po(2)$, $X_2 \in Po(2)$ och $X_3 \in Po(3)$. Låt $Y = X_1 + X_2 + X_3$. Beräkna $P(Y = 2)$.

Uppgift 4

En stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{x}{2}, & 0 \leq x \leq 2, \\ 1, & x > 2. \end{cases}$$

Bestäm standardavvikelsen $D(e^{-X})$.

A: 0.432

B: 0.242

C: 0.0585

D: 0.03

Uppgift 5

Låt X och Y vara två oberoende stokastiska variabler där $X \in N(2, 3)$, dvs $E(X) = 2$, $D(X) = 3$, och $Y \in N(4, 2)$, dvs $E(Y) = 4$, $D(Y) = 2$. Låt $Z = 2Y - X$. Beräkna $P(Z > 4)$.

A: 0.345

B: 0.468

C: 0.532

D: 0.655

Uppgift 6

Låt $X \in \text{Exp}(4)$, dvs intensiteten λ är lika med fyra. Bestäm väntevärdet $E(X^2)$.

Uppgift 7

Två stickprov från två oberoende populationer. Varje stickprov uppfattas som observationer på $N(\mu_i, \sigma_i)$, $i = 1, 2$, och där vi antar att $\sigma_1 = \sigma_2$ är okända. Från de två stickproven beräknades följande sammanfattande mått:

från stickprov 1

$$n_1 = 4 \quad \bar{x}_1 = 1007.25 \quad s_1 = 143.66,$$

från stickprov 2

$$n_2 = 4 \quad \bar{x}_2 = 817.75 \quad s_2 = 73.627.$$

Beräkna den nedre gränsen i ett 95%-igt tväsidigt konfidensintervall för $\mu_1 - \mu_2$.

A: 13.14

B: -76.677

C: 25.53

D: -8.25

Uppgift 8

Antag att X_1, \dots, X_n utgör ett stickprov på $N(\mu, \sigma)$, där σ är känd. Tyko önskar testa nollhypotesen $H_0 : \mu = 2$ mot $H_1 : \mu < 2$ med hjälp av ett lämpligt konfidensintervall för μ . Testets signifikansnivå skall bli α . Vilket av nedanstående konfidensintervall för μ ska användas?

A: $I_\mu = \left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$

B: $I_\mu = \left(\bar{x} - t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}} \right)$

C: $I_\mu = \left(-\infty, \bar{x} + \lambda_\alpha \frac{\sigma}{\sqrt{n}} \right)$

D: $I_\mu = \left(\bar{x} - \lambda_\alpha \frac{\sigma}{\sqrt{n}}, +\infty \right)$

Uppgift 9

Antag att X_1, \dots, X_n utgör ett stickprov på $N(\mu, \sigma)$. Från tjugo observationer erhöles $\bar{x} = 0.46$ samt $s = 0.43$. Ange nedre gränsen för det tvåsidiga konfidensintervallet för σ med konfidensgrad 99%.

- A: 0.158
- B: 0.302
- C: 0.316
- D: 0.415

Uppgift 10

En forskare har gjort tio försök som anses vara oberoende av varandra där sannolikheten för lyckat försök är p . Låt X stå för antalet lyckade försök. Forskaren önskar pröva $H_0 : p = 1/2$ mot $H_1 : p > 1/2$. Resultatet av de tio försöken var åtta lyckade försök. Beräkna P -värdet!

- A: 0.0107
- B: 0.044
- C: 0.100
- D: 0.0547

Uppgift 11

Ett politiskt forskningsinstitut vill undersöka om det finns ett samband mellan åldersgrupp och partipreferens (Parti A eller Parti B) inför ett val. De tillfrågade 200 personer och fick följande resultat (observerade värden):

Ålder / Parti	Parti A	Parti B	Totalt
Unga (15-35)	30	50	80
Äldre (36+)	60	60	120
Totalt	90	110	200

Vi önskar testa om det finns ett statistiskt säkerställt samband mellan ålder och partipreferens. Vi beräknar teststorheten Q och får $Q = 3.04$. Vilken slutsats kan man dra då man fått denna teststorhet?

- A: Det finns ett statistiskt säkerställt samband mellan ålder och partipreferens *både* på risknivån 5 % och risknivån 1 %, ty testets P -värde är mindre än 0.05
- B: Det finns ett statistiskt säkerställt samband mellan ålder och partipreferens *både* på risknivån 5 % och risknivån 1 %, ty testets P -värde är större än 0.05
- C: Det finns *varken* ett statistiskt säkerställt samband mellan ålder och partipreferens på risknivån 5 % eller risknivån 1 %, ty testets P -värde är mindre än 0.05
- D: Det finns *varken* ett statistiskt säkerställt samband mellan ålder och partipreferens på risknivån 5 % eller risknivån 1 %, ty testets P -värde är större än 0.05

Uppgift 12

Två oberoende undersökningar gjordes på två grupper av patienter som hade blivit vaccinerade respektive inte hade blivit vaccinerade. Femhundra vaccinerade undersöktes där fyrtionio hade blivit vinterkräksjuka. I den icke vaccinerade gruppen undersöktes sexhundra där femtioåtta hade fått vinterkräksjuka. Låt p_1 stå för andelen patienter som blivit vinterkräksjuka trots att de har vaccinerats och låt p_2 stå för andelen patienter som blivit vinterkräksjuka utan att ha vaccinerats. Vi är intresserade av parametern $p_1 - p_2$.

Bestäm medelfelet för $p_1^* - p_2^*$, där $p_{1,obs}^*$ är maximum-likelihood-skattningen av p_1 och $p_{2,obs}^*$ är maximum-likelihood-skattningen av p_2 .

A: 0.018

B: 0.0973

C: 0.1757

D: 0.419

Var god vänd!

Del II

Uppgift 13

(Fortsättning på Uppgift 3.) Låt X_1 , X_2 och X_3 vara tre oberoende stokastiska variabler sådana att $X_1 \in Po(2)$, $X_2 \in Po(2)$ och $X_3 \in Po(3)$. Låt $Y_1 = X_1 + X_3$ och $Y_2 = X_2 + X_3$. Bestäm korrelationskoefficienten $\rho(Y_1, Y_2)$. (10 p)

Uppgift 14

(Fortsättning på Uppgift 6.) Låt $X \in Exp(4)$, d v s intensiteten λ är lika med fyra.

(a) Ange funktionen $b(n)$ sådan att $E(X^n) = b(n) \cdot E(X^{n-1})$, där $n = 3, 4, 5, \dots$ (4 p)

(b) Bestäm sedan $E(X^n)$ som en funktion av n , där $n = 3, 4, 5, \dots$ (4 p)

(c) Beräkna $E(X^7)$. (2 p)

Uppgift 15

I ett avancerat växthus utförs ett experiment för att avgöra om mängden belysning påverkar hur mycket jordgubbar växer. Belysningen mäts med hjälp av ett belysningsindex och jordgubbarnas vikt mäts i gram. De första fyra erhållna observationerna följer nedan.

Belysningsindex	5	10	10	15
Jordgubbsvikt (g)	20	26	34	40

Det är rimligt att tro att det föreligger ett linjärt samband mellan variablerna. Utifrån datamaterialet ovan skattas en linjär regressionsmodell

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

där y_i = jordgubbsvikt (g) beror av x_i = belysningsindex och $\varepsilon_i \in N(0, \sigma)$ betecknar slumpmässiga fel, $i = 1, \dots, 4$, och σ är okänd. Minsta-kvadrat-skattningarna av regressionskoefficienterna α och β blev $\alpha_{obs}^* = 10$ respektive $\beta_{obs}^* = 2$.

Bestäm $I_\beta(0.95)$, dvs ett 95% konfidensintervall för β , den effekt som belysningindex har på jordgubbsvikten. Är effekten signifikant på 5% nivån? På 10% nivån? Motivera. (10 p)

Uppgift 16

Man vill undersöka om följande data kommer från en binomialfördelning. Dvs, H_0 : data kommer från $Bin(3, p)$ mot H_1 : data kommer inte från $Bin(3, p)$, där p är en okänd parameter.

Observation	0	1	2	3	Totalt
Antal	13	30	30	27	100

Pröva hypotesen H_0 mot H_1 med ett χ^2 -test (dvs med teststorheten Q). Kan man förkasta H_0 på risknivån 1%? Kan man förkasta H_0 på risknivån 5%? Motivera och sedan tolka resultatet. (10 p)

Lycka till!



Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1920/SF1921
SANNOLIKHETSTEORI OCH STATISTIK,
FREDAG 6 MARS 2026 KL 8.00–13.00.

Del I, Svar.

1. D
2. 0.01818 (0.0182 eller $\frac{1}{55}$ är OK)
3. 0.02234 (0.0223 är OK)
4. B
5. D
6. 0.125 (1/8 är OK)
7. D
8. C
9. B
10. D
11. D
12. A

Del II, Svar.

13. $\rho(Y_1, Y_2) = 0.6$
14. a) $b(n) = n/4$; b) $E(X^n) = n!/4^n$; c) $E(X^7) = 7!/4^7 = 0.3076$
15. $I_\beta(0.90) = (0.35, 3.65)$, $I_\beta(0.95) = (-0.43, 4.43)$. Därmed är effekten signifikant på 10% nivån dock ej är signifikant på 5% sign.nivån, ty noll ingår ej resp ingår i motsvarande k.i.
16. $p_{obs}^* = 0.57$, $E_i = \{8, 32, 42, 18\}$, $Q_{obs} = 11.18$ som är större än $\chi_{0.01}^2(4 - 1 - 1) = \chi_{0.01}^2(2) = 9.21$ och därför kan H_0 förkastas på risknivån 1% (och isf även på risknivån 5%).

Del I, Lösningsförslag.**Uppgift 1**

Vi har:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \text{|rita Venn-diagram|} = \frac{P(A \cup B) - P(A \cap B^*) - P(B \cap A^*)}{P(A \cup B) - P(B \cap A^*)} =$$

$$= \frac{0.9 - 0.2 - 0.4}{0.9 - 0.4} = \frac{0.3}{0.5} = \frac{3}{5} = 0.6$$

Uppgift 2

$$P(\text{exakt tre skivor gravad lax av tre på måfå tagna}) = \frac{\binom{3}{0} \binom{4}{3} \binom{5}{0}}{\binom{12}{3}} = \frac{4}{\frac{12 \cdot 11 \cdot 10}{3 \cdot 2 \cdot 1}} = \frac{1}{55} \simeq 0.01818$$

Uppgift 3

Observera först att $Y = X_1 + X_2 + X_3 \in Po(\mu_{X_1} + \mu_{X_2} + \mu_{X_3})$, där $\mu_{X_1} + \mu_{X_2} + \mu_{X_3} = 2 + 2 + 3 = 7$, ty X_1, X_2 och X_3 är oberoende s.v. Därför har vi:

$$P(Y = 2) = P(\leq 2) - P(Y \leq 1) = \text{|Tabell 5, för } \mu = 7\text{|} = 0.02964 - 0.00730 = 0.02234$$

Uppgift 4

Först, observera att täthetsfunktionen $f_X(x) = F'_X(x) = 1/2$, då $0 \leq x \leq 2$, och 0 annars. Vi vet att $D(e^{-X}) = \sqrt{V(e^{-X})}$.

Vidare, $V(e^{-X}) = E(e^{-2X}) - (E(e^{-X}))^2$, där

$$E(e^{-X}) = \int_0^2 e^{-x}(1/2)dx = \frac{1}{2}(-e^{-x})\Big|_0^2 = \frac{1}{2}(1 - e^{-2})$$

och

$$E(e^{-2X}) = \int_0^2 e^{-2x}(1/2)dx = \frac{1}{2}(-1/2)e^{-2x}\Big|_0^2 = \frac{1}{4}(1 - e^{-4}).$$

Därför har vi:

$$V(e^{-X}) = \frac{1}{4}(1 - e^{-4}) - \frac{1}{4}(1 - e^{-2})^2 = \frac{1}{4}(1 - e^{-4} - 1 + 2e^{-2} - e^{-4}) = \frac{1}{2}(e^{-2} - e^{-4})$$

och därmed

$$D(e^{-X}) = \sqrt{V(e^{-X})} = \sqrt{\frac{1}{2}(e^{-2} - e^{-4})} \simeq \sqrt{0.0585} \simeq 0.242$$

Uppgift 5

Pga oberoendet av X och Y , och eftersom dom både är normalfördelade s.v.:

$$Z \in N(2 \cdot 4 - 2, \sqrt{2^2 \cdot 2^2 + 3^2}) = N(6, 5).$$

Därmed,

$$\begin{aligned} P(Z > 4) &= 1 - P(Z \leq 4) = 1 - \Phi\left(\frac{4-6}{5}\right) = 1 - \Phi(-0.4) = |\text{symmetri av } N(0, 1) \text{ runt } 0| = \\ &= \Phi(0.4) = |\text{Tabell 1}| = 0.6554 \simeq 0.655 \end{aligned}$$

Uppgift 6

$$E(X^2) = V(X) + (E(X))^2 = |X \in \text{Exp}(4), \text{ och FS}| = \frac{1}{4^2} + \left(\frac{1}{4}\right)^2 = \frac{1}{16} + \frac{1}{16} = \frac{1}{8} = 0.125$$

Uppgift 7

Vi har enl FS (för två stickprov med okända dock samma standardavvikelse $\sigma_1 = \sigma_2 = \sigma$):

$$I_{\mu_1 - \mu_2}(0.95) = \left(\bar{x}_1 - \bar{x}_2 \pm t_{0.025}(4+4-2) \cdot s \cdot \sqrt{\frac{1}{4} + \frac{1}{4}} \right),$$

där

$$s = \sqrt{\frac{(4-1)s_1^2 + (4-1)s_2^2}{4+4-2}}.$$

Dvs, den nedre gränsen blir

$$\begin{aligned} \bar{x}_1 - \bar{x}_2 - t_{0.025}(6) \cdot \sqrt{\frac{3 \cdot s_1^2 + 3 \cdot s_2^2}{6}} \cdot \sqrt{\frac{1}{2}} &= |t_{0.025}(6) = 2.45, \text{ enl Tabell 3}| = \\ &= 1007.25 - 817.75 - (2.45) \cdot \frac{1}{2} \cdot \sqrt{(143.66)^2 + (73.627)^2} = -8.249799 \simeq -8.25 \end{aligned}$$

Uppgift 8

Eftersom vi har en ensidig mothypotes H_1 – faller de både svarsalternativen A och B . Svarsalternativet C är det korrekta, ty om nollhypotesens värde $\mu = 2 \notin I_\mu = \left(-\infty, \bar{x} + \lambda_\alpha \frac{\sigma}{\sqrt{n}}\right)$ så måste i så fall detta innebära: a) att H_0 förkastas på signifikansnivån α ; b) att $\left(-\infty, \bar{x} + \lambda_\alpha \frac{\sigma}{\sqrt{n}}\right) < 2$, där $I_\mu = I_\mu(1 - \alpha) = \left(-\infty, \bar{x} + \lambda_\alpha \frac{\sigma}{\sqrt{n}}\right)$ representerar μ med konfidensgraden $1 - \alpha$, dvs, $\mu < 2$.

Uppgift 9

Vi har enl FS:

$$I_{\sigma}(0.99) = \left(s \sqrt{\frac{n-1}{\chi_{\frac{0.01}{2}}^2(n-1)}}, s \sqrt{\frac{n-1}{\chi_{1-\frac{0.01}{2}}^2(n-1)}} \right),$$

där $n = 20$, $s = 0.43$. Dvs, den nedre gränsen blir

$$\begin{aligned} 0.43 \sqrt{\frac{19}{\chi_{0.005}^2(19)}} &= |\chi_{0.005}^2(19) = 38.6, \text{ enl Tabell 4}| = \\ &= 0.43 \sqrt{\frac{19}{38.6}} = 0.30168 \simeq 0.302 \end{aligned}$$

Uppgift 10

Noll- och mothypoteser är:

$$H_0 : p = 0.5$$

$$H_1 : p > 0.5$$

Vi har därför att

$$\begin{aligned} P\text{-värde} &= P(\text{att få det vi fått eller ännu extremare} | H_0 \text{ sann}) = \\ &= P(X \geq 8 | X \in \text{Bin}(10, 0.5)) = |\text{komplementet och } X \text{ är diskret s.v.}| = \\ &= 1 - P(X \leq 7 | X \in \text{Bin}(10, 0.5)) = |\text{enl Tabell 6 för } n = 10, p = 0.5| = \\ &= 1 - 0.94531 = 0.05469 \simeq 0.0547 \end{aligned}$$

Uppgift 11

Oberoendetest. Noll- och mothypoteser är:

H_0 : Det finns inget statistiskt säkerställt samband mellan ålder och partiprefenser.

H_1 : Det finns ett statistiskt säkerställt samband mellan ålder och partiprefenser.

De två kritiska χ^2 -värden är enl Tabell 4:

$$\begin{aligned} 1) \chi_{0.05}^2((r-1)(s-1)) &= \chi_{0.05}^2((2-1)(2-1)) = \chi_{0.05}^2(1) = 3.84, \\ 2) \chi_{0.01}^2((r-1)(s-1)) &= \chi_{0.01}^2((2-1)(2-1)) = \chi_{0.01}^2(1) = 6.63. \end{aligned}$$

Eftersom $Q = Q_{obs} = 3.04 < 3.84 = \chi_{0.05}^2(1)$ och $Q = Q_{obs} = 3.04 < 6.63 = \chi_{0.01}^2(1)$ så går det varken att förkasta H_0 på risknivån 5% eller på risknivån 1%. Därmed, både alternativ A och B faller. Vidare, eftersom $Q = 3.04 < 3.84 = \chi_{0.05}^2(1)$ så är P -värdet > 0.05 . Därmed faller även alternativ C.

Uppgift 12

Man kan visa att maximum-likelihood-skattningen av p_1 är

$$p_{1,obs}^* = \frac{x_1}{n_1},$$

där $n_1 = 500$, och $x_1 = 49$ är en observation av en s.v. $X_1 \in Bin(500, p_1)$. Detta genom att:

1) ställa upp likelihood-funktionen

$$L(p_1) = L(p_1; x_1) = p_{X_1}(x_1) = P(X_1 = x_1) = \binom{n_1}{x_1} p_1^{x_1} (1 - p_1)^{n_1 - x_1},$$

2) sedan logaritmera den,

$$\ln L(p_1) = \ln \binom{n_1}{x_1} + x_1 \ln p_1 + (n_1 - x_1) \ln(1 - p_1),$$

3) sedan derivera $\ln L(p_1)$ med avseende på p_1 och lösa sen ekvation

$$\frac{d}{dp_1} (\ln L(p_1)) = \frac{x_1}{p_1} - \frac{n_1 - x_1}{1 - p_1} = 0,$$

där lösningen map p_1 blir $\frac{x_1}{n_1} = p_{1,obs}^*$, som är då ML-skattningen av p_1 .

På samma sätt kan man visa att maximum-likelihood-skattningen av p_2 är

$$p_{2,obs}^* = \frac{x_2}{n_2},$$

där $n_2 = 600$, och $x_2 = 58$ är en observation av en s.v. $X_2 \in Bin(600, p_2)$. Observera att X_1 och X_2 är oberoende s.v. Vidare beräknar man först variansen för $p_1^* - p_2^* = X_1/n_1 - X_2/n_2$:

$$\begin{aligned} V(p_1^* - p_2^*) &= V\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = |\text{oberoende undersökningar}| = V\left(\frac{X_1}{n_1}\right) + V\left(\frac{X_2}{n_2}\right) = \\ &= \frac{V(X_1)}{n_1^2} + \frac{V(X_2)}{n_2^2} = |X_1, X_2 \text{ är Bin-fördelade}| = \frac{n_1 p_1 (1 - p_1)}{n_1^2} + \frac{n_2 p_2 (1 - p_2)}{n_2^2} = \\ &= \frac{p_1 (1 - p_1)}{n_1} + \frac{p_2 (1 - p_2)}{n_2}. \end{aligned}$$

Därmed, blir medelfelet för $p_1^* - p_2^*$:

$$\begin{aligned} d(p_1^* - p_2^*) &= (D(p_1^* - p_2^*))_{obs}^* = \left(\sqrt{V(p_1^* - p_2^*)}\right)_{obs}^* = \left(\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)_{obs}^* \\ &= \sqrt{\frac{p_{1,obs}^*(1-p_{1,obs}^*)}{n_1} + \frac{p_{2,obs}^*(1-p_{2,obs}^*)}{n_2}} = \sqrt{\frac{49}{500} \cdot \left(1 - \frac{49}{500}\right) + \frac{58}{600} \cdot \left(1 - \frac{58}{600}\right)} \simeq 0.018 \end{aligned}$$

Del II, Lösningsförslag.**Uppgift 13**

Vi har enl FS:

$$\rho(Y_1, Y_2) = \frac{C(Y_1, Y_2)}{D(Y_1)D(Y_2)}.$$

Vidare,

$$\begin{aligned} C(Y_1, Y_2) &= C(X_1 + X_3, X_2 + X_3) = \\ &= C(X_1, X_2 + X_3) + C(X_3, X_2 + X_3) = C(X_2 + X_3, X_1) + C(X_2 + X_3, X_3) = \\ &= C(X_2, X_1) + C(X_3, X_1) + C(X_2, X_3) + C(X_3, X_3) = C(X_3, X_3) = V(X_3) = 3, \end{aligned}$$

enligt FS, och ty X_1, X_2, X_3 är oberoende och därför är dom parvis okorrelerade också, dvs $C(X_2, X_1) = C(X_1, X_2) = 0$, $C(X_3, X_1) = C(X_1, X_3) = 0$ och $C(X_2, X_3) = 0$.

Också, vi har:

$$\begin{aligned} D(Y_1) &= \sqrt{V(Y_1)} = \sqrt{V(X_1 + X_3)} = |X_1 \text{ och } X_3 \text{ är oberoende}| = \\ &= \sqrt{V(X_1) + V(X_3)} = \sqrt{2 + 3}. \end{aligned}$$

På samma sätt,

$$D(Y_2) = \sqrt{V(X_2 + X_3)} = \sqrt{V(X_2) + V(X_3)} = \sqrt{2 + 3}.$$

Därmed,

$$\rho(Y_1, Y_2) = \frac{C(Y_1, Y_2)}{D(Y_1)D(Y_2)} = \frac{3}{\sqrt{5} \cdot \sqrt{5}} = \frac{3}{5} = 0.6$$

Uppgift 14

(a) För $X \in \text{Exp}(\lambda)$, $\lambda > 0$, gäller att

$$\begin{aligned} E(X^n) &= \int_0^\infty x^n \lambda e^{-\lambda x} dx = |\text{partiell integration, } x^n = u, e^{-\lambda x} dx = dv| = \\ &= \lambda \left(\left(-\frac{1}{\lambda} x^n e^{-\lambda x} \right)_0^\infty + \frac{n}{\lambda} \int_0^\infty x^{n-1} e^{-\lambda x} dx \right) = n \int_0^\infty x^{n-1} e^{-\lambda x} dx = \\ &= \frac{n}{\lambda} \int_0^\infty x^{n-1} \lambda e^{-\lambda x} dx = \frac{n}{\lambda} E(X^{n-1}), \end{aligned}$$

eftersom $\left(-\frac{1}{\lambda} x^n e^{-\lambda x} \right)_0^\infty = 0$ även då $x \rightarrow \infty$ (enl L'hospital regel, om vi deriverar n -gångar både täljaren (x^n) och nämnaren ($e^{-\lambda x}$)).

Därmed,

$$b(n) = \frac{n}{\lambda} = \frac{n}{4},$$

för $X \in \text{Exp}(4)$.

(b) Från (a),

$$\begin{aligned}
E(X^n) &= \frac{n}{4} E(X^{n-1}) = |\text{iterativt}| = \\
&= \frac{n}{4} \cdot \frac{n-1}{4} \cdot E(X^{n-2}) = \dots = \frac{n(n-1)(n-2) \cdot \dots \cdot (n-(n-3))}{4^{(n-3)+1}} \cdot E(X^{n-(n-3)-1}) = \\
&= \frac{n(n-1)(n-2) \cdot \dots \cdot 4 \cdot 3}{4^{n-2}} \cdot E(X^2) = |E(X^2) = 1/8, \text{ enl Uppgift 3}| = \\
&= \frac{n(n-1)(n-2) \cdot \dots \cdot 4 \cdot 3}{4^{n-2}} \cdot \frac{1}{4 \cdot 2} = \frac{n(n-1)(n-2) \cdot \dots \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4^{n-2}} \cdot \frac{1}{4 \cdot 2 \cdot 2} = \frac{n!}{4^n},
\end{aligned}$$

där $n! = n(n-1)(n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$.

(c) Från (b),

$$E(X^7) = \frac{7!}{4^7} = \frac{7 \cdot 6 \cdot 5 \cdot \dots \cdot 3 \cdot 2 \cdot 1}{4^7} \simeq 0.3076$$

Uppgift 15

Vi testar följande H_0 och H_1 : $H_0 : \beta = 0$ (dvs, effekten i frågan **är ej** signifikant) $H_1 : \beta \neq 0$ (dvs, effekten i frågan **är** signifikant)Observera att $n = 4$, $\alpha_{obs}^* = 10$, $\beta_{obs}^* = 2$. Vi har enl FS:

$$I_\beta(1 - \alpha) = \left(\beta_{obs}^* \pm t_{\alpha/2}(4 - 2) \frac{s}{\sqrt{S_{xx}}} \right),$$

där

$$\begin{aligned}
s &= \sqrt{\frac{1}{4-2} \sum_{i=1}^4 (y_i - (\alpha_{obs}^* + \beta_{obs}^* x_i))^2} = \\
&= \sqrt{\frac{1}{2} \left((20 - (10 + 2 \cdot 5))^2 + (26 - (10 + 2 \cdot 10))^2 + (34 - (10 + 2 \cdot 10))^2 + (40 - (10 + 2 \cdot 15))^2 \right)} \\
&= 4
\end{aligned}$$

och

$$S_{xx} = \sum_{i=1}^4 (x_i - \bar{x})^2 = \sum_{i=1}^4 x_i^2 - n(\bar{x})^2 = (5^2 + 10^2 + 10^2 + 15^2 - 4 \cdot (10)^2) = 50,$$

och, från Tabell 3, $t_{0.025}(2) = 4.30$ respektive $t_{0.05}(2) = 2.92$. Därmed,

$$I_\beta(0.95) = \left(\beta_{obs}^* \pm t_{0.025}(2) \frac{s}{\sqrt{S_{xx}}} \right) = \left(2 \pm 4.3 \cdot \frac{4}{\sqrt{50}} \right) = \left(2 \pm 2.43245 \right) \simeq (-0.43; 4.43)$$

och

$$I_\beta(0.9) = \left(\beta_{obs}^* \pm t_{0.05}(2) \frac{s}{\sqrt{S_{xx}}} \right) = \left(2 \pm 2.92 \cdot \frac{4}{\sqrt{50}} \right) = \left(2 \pm 1.652 \right) \simeq (0.35; 3.65).$$

Slutsats:

- 1) 5%-sign.nivån: $0 \in I_\beta(0.95) = (-0.43; 4.43)$ och därför kan man inte förkasta $H_0 : \beta = 0$, och därmed är effekten (som belysningindex har på jordgubbsvikten) **inte signifikant** på 5% nivån;
- 2) 10%-sign.nivån: $0 \notin I_\beta(0.9) = (0.35; 3.65)$ och därför kan man förkasta $H_0 : \beta = 0$, och därmed är effekten **signifikant** på 10% nivån.

Uppgift 16

Först skattar vi den okända parametern p från datamaterialet. Vi använder oss av ML-skattningen av p . I tabellen har vi 100 st observationer x_1, x_2, \dots, x_{100} på s.v. X_1, X_2, \dots, X_{100} som samtliga är oberoende av varandra och alla är $Bin(3, p)$ -fördelade. Från Uppgift 12 vet vi att ML-skattningen av p är $p_{obs,ML}^* = \frac{x_1}{3}$ om vi har endast en observation x_1 som kommer från s.v. $X_1 \in Bin(3, p)$. Vi utvecklar detta och tittar på två s.v. nu. Låt oss nu alltså ha två observationer x_1 och x_2 på oberoende s.v. $X_1 \in Bin(n_1, p)$ resp $X_2 \in Bin(n_2, p)$. Vad är ML-skattningen av p då? På samma sätt som i Uppgift 12:

- 1) ställer upp först likelihood-funktionen

$$\begin{aligned} L(p) &= L(p; x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2) = \\ &= \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \binom{n_2}{x_2} p^{x_2} (1-p)^{n_2-x_2}, \end{aligned}$$

- 2) sedan logaritmera den,

$$\ln L(p) = \ln \binom{n_1}{x_1} + x_1 \ln p + (n_1 - x_1) \ln(1-p) + \ln \binom{n_2}{x_2} + x_2 \ln p + (n_2 - x_2) \ln(1-p),$$

- 3) och sedan derivera $\ln L(p)$ med avseende på p , sätta den till 0 och lösa ekvationen:

$$\frac{d}{dp} (\ln L(p)) = \frac{x_1}{p} - \frac{n_1 - x_1}{1-p} + \frac{x_2}{p} - \frac{n_2 - x_2}{1-p} = 0,$$

där lösningen map p blir $\frac{x_1+x_2}{n_1+n_2} = p_{obs,ML}^*$, som är ML-skattningen av p då. I vårt fall blir alltså ML-skattningen av p lika med $p_{obs}^* = p_{obs,ML}^* = \frac{x_1+x_2}{3+3} = \frac{x_1+x_2}{3 \cdot 2} = \frac{x_1+x_2}{6}$, om vi har endast två observationer x_1 och x_2 som kommer från oberoende s.v. $X_1 \in Bin(3, p)$ resp $X_2 \in Bin(3, p)$.

På samma sätt, eftersom i denna uppgift har vi 100 st observationer x_1, x_2, \dots, x_{100} på s.v. X_1, X_2, \dots, X_{100} som är samtliga oberoende och alla är $Bin(3, p)$ -fördelade, så blir ML-skattningen av p lika med:

$$\begin{aligned} p_{obs}^* &= p_{obs,ML}^* = \frac{x_1 + x_2 + \dots + x_{100}}{3 + 3 + \dots + 3} = \frac{x_1 + x_2 + \dots + x_{100}}{3 \cdot 100} = \frac{x_1 + x_2 + \dots + x_{100}}{300} = \\ &= |\text{med användning till vårt datamaterial}| = \frac{0 \cdot 13 + 1 \cdot 30 + 2 \cdot 30 + 3 \cdot 27}{300} \simeq 0.57 \end{aligned}$$

DVS, våran skattning $p_{obs}^* = 0.57$ är det bästa tänkbara värdet på den okända parametern p (med tanke på ML-metoden).

Detta är χ^2 -test av given fördelning med skattade parametrar, dvs $k = 1$ parameter p är skattade från det angivna datamaterialet. DVS, man vill undersöka nu om det givna datamaterialet kommer från en binomialfördelning med parametrar $n = 3$ och $p = 0.57$. Vi har,

H_0 : data kommer från $Bin(3, 0.57)$,

mot

H_1 : data kommer inte från $Bin(3, 0.57)$.

Signifikansnivån (eller risknivån) är $\alpha = 0.01$.

Vi har för s.v. $X \in Bin(3, 0.57)$:

$$P(X = 0) = \binom{3}{0} 0.57^0 0.43^3 = 0.43^3 \simeq 0.08,$$

$$P(X = 1) = \binom{3}{1} 0.57^1 0.43^2 = 3(0.57^1) 0.43^2 \simeq 0.32,$$

$$P(X = 2) = \binom{3}{2} 0.57^2 0.43^1 = 3(0.57^2) 0.43^1 \simeq 0.42,$$

$$P(X = 3) = \binom{3}{3} 0.57^3 0.43^0 = 0.57^3 \simeq 0.18$$

Därmed, *förväntat antal obs.*, E_i , under H_0 är angivet (tillsammans med det ursprungliga datat, O_i) i tabellen nedan:

Observation	0	1	2	3	Totalt
Antal, O_i	13	30	30	27	100
Förväntat antal, E_i	8	32	42	18	100

där $8 = 100 \cdot P(X = 0)$, $32 = 100 \cdot P(X = 1)$, $42 = 100 \cdot P(X = 2)$, $18 = 100 \cdot P(X = 3)$ (observera att alla de fyra förväntade antal är större än 5, vilket måste uppfyllas för att χ^2 -testet ska vara tillförlitligt).

Teststorheten Q under H_0 beräknas till

$$Q_{obs} = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{(13 - 8)^2}{8} + \frac{(30 - 32)^2}{32} + \frac{(30 - 42)^2}{42} + \frac{(27 - 18)^2}{18} \simeq$$

$$\simeq 3.125 + 0.125 + 3.43 + 4.5 \simeq 11.18$$

Från Tabell 4, med tanke på att $k = 1$ parameter har blivit skattat (dvs, den okända parametern p) är det kritiska värdet lika med

$$\chi_{\alpha}^2(r - 1 - k) = \chi_{0.01}^2(4 - 1 - 1) = \chi_{0.01}^2(2) = 9.21,$$

och därmed **kan nollhypotesen H_0 förkastas** på risknivån 1%, ty $Q_{obs} \simeq 11.18 > 9.21 = \chi_{\alpha}^2(r - 1 - 1)$, där $r = 4$ är antal grupper i datamaterialet.

Slutsats och tolkning: På risknivån 1% kan vi förkasta nollhypotesen H_0 , dvs vi kan påstå att det angivna datamaterialet inte kan tänka sig komma från en binomialfördelning ($Bin(3, p)$ i det här fallet) på 1% signifikansnivån. Och därmed samma slutsats gäller även på 5% signifikansnivån (ty, P -värdet $< 0.01 < 0.05$, detta eftersom H_0 går att förkasta på risknivån 1%).