



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1914/SF1915/SF1916 SANNOLIKHETSTEORI OCH STATISTIK,
ONSDAG 17 DECEMBER 2025 KL 8.00–13.00.

Examinator SF1914/SF1915/SF1916: Mykola Shykula, 08-790 66 44.

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. 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"). Gränsen för godkänt är 9 poäng av 12 möjliga. 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.

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

En underleverantör till en kameratillverkare gör skärmar till digitalkameror. 7 % av skärmarna blir defekta vid tillverkningen. Vid den efterföljande kvalitetskontrollen blir alla felfria skärmar klassade som felfria, men 5 % av de defekta skärmarna slinker igenom kontrollen och klassas som felfria. De skärmar som klassas som defekta kasseras. Beräkna hur stor andel (i %) av de skärmar som klassas vid kvalitetskontrollen som felfria kommer i själva verket att bli defekta?

A: ca 6.67 %

B: ca 0.37 %

C: 93 %

D: 100 %

Uppgift 2

Låt $X \in \text{Exp}(2)$, dvs den stokastiska variabeln X har täthetsfunktionen

$$f_X(x) = \begin{cases} 2e^{-2x}, & x \geq 0, \\ 0, & \text{annars.} \end{cases}$$

Bestäm väntevärdet $E(-2X^2 + X - \frac{1}{2})$.

A: $-\frac{3}{2}$

B: $\frac{1}{2}$

C: 0

D: -1

Uppgift 3

De två diskreta stokastiska variablerna X och Y är sådana att

$$\begin{aligned} P(X=0, Y=0) &= 1/12, & P(X=0, Y=1) &= 1/6, \\ P(X=1, Y=0) &= 1/4, & P(X=1, Y=1) &= 1/2. \end{aligned}$$

Således antar X värdena 0 och 1, samt Y antar också värdena 0 och 1.

Bestäm den betingade sannolikheten $P(X+Y=1 \mid Y=0)$.

A: 5/12

B: 2/3

C: 3/4

D: 1

Uppgift 4

Låt X vara en stokastisk variabel sådan att $X \in \text{Bin}(13, 0.8)$.

Beräkna $P(X \geq 10)$.

A: ca 0.75

B: ca 0.5

C: ca 0.25

D: ca 0.05

Uppgift 5

Låt X_1 och X_2 vara oberoende stokastiska variabler sådana att $X_1 \in N(-1, 1)$, $X_2 \in N(3, 3)$, d.v.s. $E(X_1) = -1$, $V(X_1) = 1$, och $E(X_2) = 3$, $V(X_2) = 9$. Låt vidare $Y = X_2 + 3X_1$.

Bestäm a så att $P(Y \leq a) = 0.05$.

- A: 0
- B: -1.645
- C: -5.7
- D: -6.98

Uppgift 6

En stokastisk variabel X har följande täthetsfunktion

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

Bestäm täthetsfunktionen för $Y = X^2$ i punkten $y = 2$, dvs bestäm $f_Y(2)$.

- A: 0.1465
- B: 0.293
- C: 1.414
- D: 0.172

Uppgift 7

Låt x_i , $i = 1, 2, 3$, vara observationer av de oberoende stokastiska variablerna X_i , $i = 1, 2, 3$, vilka alla är gammafördelade $\Gamma(2, \lambda)$, dvs de har täthetsfunktionen

$$f_{X_i}(x) = \begin{cases} \frac{\lambda^2}{1!} x e^{-\lambda x}, & \text{om } x \geq 0, \\ 0, & \text{annars,} \end{cases}$$

för $i = 1, 2, 3$, och $\lambda > 0$. Detta medför att

$$E(X_i) = \frac{2}{\lambda} \quad \text{och} \quad V(X_i) = \frac{2}{\lambda^2}, \quad i = 1, 2, 3.$$

Bestäm MK-skattningen av λ om det är följande värden som har erhållits: $x_1 = 1.22$, $x_2 = 0.54$, och $x_3 = 1.41$

- A: 1.893
- B: 1.376
- C: 1.057
- D: 0.946

Uppgift 8

Låt $X \in N(\mu_X, \sigma_X)$ och $Y \in N(\mu_Y, \sigma_Y)$. Antag att $\sigma_X = \sigma_Y$. Vi gör 10 observationer på X och får att $\bar{x} = 19.5$ och $s_X^2 = 4.06$. Vi gör även 10 observationer på Y och får att $\bar{y} = 13.9$ och $s_Y^2 = 3.48$. Vi antar att alla observationer är utfall av stokastiska variabler som alla är oberoende. Ta fram ett nedåt begränsat konfidensintervall för skillnaden $\mu_X - \mu_Y$ som har konfidensgrad 90 %.

A: $(5.02, \infty)$ B: $(4.85, \infty)$ C: $(4.45, \infty)$ D: $(4.10, \infty)$ **Uppgift 9**

$x_1, \dots, x_5$ är oberoende observationer på en $N(\mu, \sigma)$ -fördelning, där μ och σ är okända. Stickprovsmedelvärdet blev $\bar{x} = 0.16$ och stickprovsstandardavvikelsen blev $s = 0.0158$. Ange övre gränsen i det ensidigt uppåt begränsade konfidensintervallet för σ som har konfidensgraden 99 %.

A: 0.058

B: 0.069

C: 0.459

D: 0.549

Uppgift 10

Antag att $X \in Po(\mu)$. Vi vill testa $H_0 : \mu = 10$ mot alternativet $H_1 : \mu = 15$. Vi förkastar H_0 till förmån för H_1 om vi får ett utfall $x \geq 15$. Vi får en observation: $x = 13$.

Bestäm testets styrka.

A: 0.236

B: 0.432

C: 0.534

D: 0.732

Uppgift 11

En forskargrupp vill testa H_0 mot H_1 . Forskargruppen består av 8 personer. Var och en av dem tillämpar ett test med 10 % signifikansnivå och 75 % styrka. Testen baseras på ett stickprov som består av 10 observationer.

Antag att H_0 är sann. Hur stor är sannolikheten att minst en av de åtta forskarna (felaktigt) drar slutsatsen att H_0 är falsk.

A: 0.10

B: 0.19

C: 0.25

D: 0.57

Uppgift 12

Följande datamaterial beskriver hur försäljningen av en vara beror av hur mycket som spenderats på reklam.

Reklamkostnad (kkkr)	22	29	31	35	42
Försäljning (kkkr)	105	108	107	100	89

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 = försäljning (kkkr) beror av x_i = reklamkostnad (kkkr) och ε_i betecknar slumpmässiga fel. Bestäm minsta-kvadrat-skattningarna α_{obs}^* och β_{obs}^* av regressionskoefficienterna α resp β .

A: $\alpha_{obs}^* = 129.3$, $\beta_{obs}^* = 0.865$

B: $\alpha_{obs}^* = 129.3$, $\beta_{obs}^* = -0.865$

C: $\alpha_{obs}^* = 74.3$, $\beta_{obs}^* = 0.865$

D: $\alpha_{obs}^* = 74.3$, $\beta_{obs}^* = -0.865$

Var god vänd!

Del II

Uppgift 13

Livslängden för två olika typer av elektroniska komponenter beskrivs av de stokastiska variablerna X och Y som är oberoende och exponentialfördelade med täthetsfunktionerna

$$f_X(x) = \begin{cases} \frac{1}{2}e^{-x/2}, & \text{om } x \geq 0, \\ 0, & \text{annars,} \end{cases}$$

respektive

$$f_Y(y) = \begin{cases} \frac{1}{3}e^{-y/3}, & \text{om } y \geq 0, \\ 0, & \text{annars.} \end{cases}$$

Beräkna sannolikheten $P(X > Y)$.

Ledning: integrera den simultana täthetsfunktionen över ett lämpligt område. (10 p)

Uppgift 14

(Fortsättning på Uppg 8 från oktobertentan.) Emma undrar om utomhustemperaturen är annorlunda i Södertälje än på Liljeholmstorget. För att undersöka detta mäter hon temperaturen i Celsius under fem olika dagar i juli. Mätningarna under olika dagar är oberoende av varandra.

Tid	08:15, Dag 1	10:00, Dag 2	12:00, Dag 3	16:00, Dag 4	18:00, Dag 5
Södertälje	16.5	18.3	22.6	21.2	19.0
Liljeholmstorget	15.3	17.7	22.9	19.5	18.1

Ange det ungefärliga P -värdet för ett hypotestest för att testa om utomhustemperaturen i Södertälje är signifikant högre än på Liljeholmstorget. Svaret kan även ges i form av ett intervall. Tolka resultatet. Antag också att temperaturskillnaderna är normalfördelade. (10 p)

Uppgift 15

En stor nationell studie av mat- och träningsvanor hos studenter har visat att 30 % av studenterna tränar tre gånger i veckan eller mer medan 50 % tränar en till två gånger och 20 % tränar inte alls. En motsvarande undersökning gjordes i ett slumpmässigt urval av 150 studenter i *StorStad*, som visade att 56 av dessa tränar tre gånger i veckan eller mer, 86 tränar en till två gånger i veckan och resten tränar inte.

Utför ett statistiskt test på 5 % signifikansnivå för att avgöra om studenterna i *StorStad* tränar mer än vad som är vanligt för studenter. (10 p)

Uppgift 16

Visa, med hjälp av kunskaper i sannolikhetslära och en lämplig approximation, att

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}.$$

(10 p)

Lycka till!



Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1914/SF1915/SF1916
SANNOLIKHETSTEORI OCH STATISTIK,
ONSDAG 17 DECEMBER 2025 KL 8.00–13.00.

Del I, Svar.

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

Del I, Lösningsförslag.**Uppgift 1**

DT=Defekt vid Tillverkningen, FT=Felfri vid Tillverkningen, KF=Klasas som Felfria. Vi har $P(DT) = 0.07$, $P(FT) = 0.93 = P(DT^*)$, och även de betingade sannolikheterna, $P(KF|DT) = 0.05$, $P(KF|FT) = 1.0$

Vi söker $P(DT|KF)$. Från Bayes sats,

$$\begin{aligned} P(DT|KF) &= \frac{P(KF \cap DT)}{P(KF)} = |\text{Betingade sannolikheten och Lagen om total sannolikhet}| = \\ &= \frac{P(KF|DT) P(DT)}{P(KF|DT) P(DT) + P(KF|FT) P(FT)} = \frac{(0.05)(0.07)}{(0.05)(0.07) + (1.0)(0.93)} = 0.003749 (\approx 0.37\%) \end{aligned}$$

Uppgift 2

Vi har:

$$\begin{aligned} E\left(-2X^2 + X - \frac{1}{2}\right) &= -2E(X^2) + E(X) - 1/2 = |X \in \text{Exp}(2), \text{dvs}, E(X) = 1/2| = \\ &= -2E(X^2) = -2\left(V(X) + (E(X))^2\right) = |X \in \text{Exp}(2)| = \\ &= -2\left(1/(2^2) + (1/2)^2\right) = -2(1/4 + 1/4) = -1 \end{aligned}$$

Uppgift 3

Vi har:

$$\begin{aligned} P(X + Y = 1 \mid Y = 0) &= \frac{P(X + Y = 1, Y = 0)}{P(Y = 0)} = \frac{P(X = 1, Y = 0)}{P(Y = 0)} = \\ &= \frac{\frac{1}{4}}{\frac{1}{12} + \frac{1}{4}} = \frac{\frac{3}{12}}{\frac{4}{12}} = \frac{3}{4} \end{aligned}$$

Uppgift 4

Låt $Y \in \text{Bin}(13, 0.2)$. Därmed,

$$P(X \geq 10) = P(Y \leq 3) = |\text{Tabell 6, för } \text{Bin}(13, 0.2)| = 0.747 \simeq 0.75$$

Uppgift 5

Vi har:

$$Y \in N(3 + 3(-1), \sqrt{9 + 3^2(1)}) = N(0, 3\sqrt{2}).$$

Vidare,

$$0.05 = P(Y \leq a) = \Phi\left(\frac{a - 0}{3\sqrt{2}}\right).$$

Därmed, från Tabell 2,

$$\frac{a - 0}{3\sqrt{2}} = -1.6449,$$

som ger

$$a = -(1.6449) \cdot 3 \cdot \sqrt{2} = -6.98$$

Uppgift 6

Observera först, att fördelningsfunktionen för s.v. X är $F_X(x) = (x - 1)^2$, då $1 \leq x \leq 2$. Vidare, fördelningsfunktionen för s.v. Y ,

$$\begin{aligned} F_Y(y) &= P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = |1 \leq x \leq 2| = P(1 \leq X \leq \sqrt{y}) = \\ &= F_X(\sqrt{y}) = (\sqrt{y} - 1)^2 = y - 2\sqrt{y} + 1, \end{aligned}$$

där $1 \leq y \leq 4$. Vi söker $f_Y(2)$, och vi har:

$$f_Y(y) = F'_Y(y) = \frac{d}{dy}(y - 2\sqrt{y} + 1) = 1 - 2\frac{1}{2\sqrt{y}} = 1 - \frac{1}{\sqrt{y}}.$$

Därmed, $f_Y(2) = 1 - \frac{1}{\sqrt{2}} \simeq 0.293$

Uppgift 7

Vi har:

$$Q = \left(x_1 - \frac{2}{\lambda}\right)^2 + \left(x_2 - \frac{2}{\lambda}\right)^2 + \left(x_3 - \frac{2}{\lambda}\right)^2.$$

Vidare, $dQ/d\lambda = 0$ ger:

$$2\left(x_1 - \frac{2}{\lambda}\right)(-2)(-1)(1/\lambda^2) + 2\left(x_2 - \frac{2}{\lambda}\right)(-2)(-1)(1/\lambda^2) + 2\left(x_3 - \frac{2}{\lambda}\right)(-2)(-1)(1/\lambda^2) = 0,$$

som ger

$$\frac{x_1 + x_2 + x_3}{3} = \frac{2}{\lambda}.$$

Därmed,

$$\lambda_{MK,obs}^* = \frac{2}{(x_1 + x_2 + x_3)/3} = \frac{2}{(1.22 + 0.54 + 1.41)/3} \simeq 1.893$$

Uppgift 8

Två oberoende populationer. Ett nedåt begränsat 90%-konfidsintervall för $\mu_X - \mu_Y$ då $\sigma_X = \sigma_Y$:

$$\begin{aligned} I_{\mu_X - \mu_Y}(0.9) &= \left(\bar{x} - \bar{y} - t_{0.1}(10 + 10 - 2) \sqrt{\frac{9s_X^2 + 9s_Y^2}{10 + 10 - 2}} \sqrt{\frac{1}{10} + \frac{1}{10}}, \infty \right) = \\ &= |t_{0.1}(10 + 10 - 2) = t_{0.1}(18) = 1.33, \text{ Tabell 3}| = \\ &= \left(19.5 - 13.9 - 1.33 \sqrt{\frac{9 \cdot (4.06) + 9 \cdot (3.48)}{18}} \sqrt{\frac{2}{10}}, \infty \right) \approx (4.45, \infty) \end{aligned}$$

Uppgift 9

Övre gränsen i det ensidigt uppåt begränsade 99%-k.i. för σ är:

$$s \sqrt{\frac{n-1}{\chi_{1-\alpha}^2(n-1)}} = (0.0158) \sqrt{\frac{4}{\chi_{0.99}^2(4)}} = |\text{Tabell 4}| = (0.0158) \sqrt{\frac{4}{0.3}} \simeq 0.058$$

Uppgift 10

Vi har:

$$H_0 : \mu = 10,$$

$$H_1 : \mu = 15.$$

Testets styrka blir:

$$\begin{aligned} h(15) &= P(\text{förkasta } H_0 \mid H_1 \text{ är sann}) = P(X \geq 15 \mid \mu = 15) = P(X \geq 15 \mid X \in Po(15)) = \\ &= 1 - P(X \leq 14 \mid X \in Po(15)) = |\text{Tabell 5 för } \mu = 15| = 1 - 0.46565 \simeq 0.534 \end{aligned}$$

Uppgift 11

Observera först att:

$$P(\text{förkasta } H_0 \mid H_0 \text{ är sann}) = \alpha = 0.1$$

Låt s.v. $X \in \text{Bin}(8, 0.1)$. Därmed söker vi $P(X \geq 1)$.

Vi har:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = |\text{Tabell 6 för } \text{Bin}(8, 0.1), x = 0| = \\ &= 1 - 0.43047 \simeq 0.57 \end{aligned}$$

Uppgift 12

Alternativ 1. Vi har:

$$\begin{aligned}\beta_{obs}^* &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2} = \\ &= \frac{(22 \cdot 105 + 29 \cdot 108 + \dots + 42 \cdot 89) - 5 \cdot ((22 + \dots + 42)/5)((105 + \dots + 89)/5)}{(22^2 + \dots + 42^2) - 5 \cdot ((22 + \dots + 42)/5)^2} \approx -0.865,\end{aligned}$$

där $n = 5$, x_i är reklamkostnad (kk), och y_i är försäljning (kk) som beror linjärt av x_i . Vidare,

$$\alpha_{obs}^* = \bar{y} - \beta_{obs}^* \cdot \bar{x} \approx 129.3$$

Alternativ 2. Ett annat sätt att förstå vilket svarsalternativ som gäller här - är att rita (på ett ungefärligt sätt) datamaterialet. Då inser man genast att lutningen β_{obs}^* på den linjära anpassningen är negativ. Och eftersom samtliga y_i ligger runt ca 100 så måste α_{obs}^* vara iaf större än 100. Därmed, är B det korrekta svarsalternativet.

Del II, Lösningförslag.**Uppgift 13**

Vi använder oss av informationen att s.v.:er X och Y är oberoende, dvs att den simultana täthetsfunktionen $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, och vi får:

$$\begin{aligned}P(X > Y) &= \iint_{\mathbb{R}^2 \cap \{(x,y) \in \mathbb{R}^2: x \geq 0, y \geq 0, y < x\}} f_{X,Y}(x, y) dx dy = |\text{oberoendet}| = \\ &= \iint_{\mathbb{R}^2 \cap \{(x,y) \in \mathbb{R}^2: x \geq 0, y \geq 0, y < x\}} f_X(x)f_Y(y) dx dy = \\ &= \int_0^\infty dx \int_0^x dy \left(\frac{1}{2} e^{-x/2} \frac{1}{3} e^{-y/3} \right) = \\ &= \int_0^\infty dx \left(\frac{1}{2} e^{-x/2} \right) \left(1 - e^{-x/3} \right) = \\ &= \int_0^\infty \left(\frac{1}{2} e^{-x/2} - \frac{1}{2} e^{-(1/2+1/3)x} \right) dx = \\ &= \left(-e^{-x/2} + \frac{1/2}{1/2+1/3} e^{-(1/2+1/3)x} \right) \Big|_{x=0}^{x=\infty} = \\ &= 0 - (-1) + 0 - \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3}} = \\ &= 1 - \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3}} = \frac{\frac{1}{3}}{\frac{1}{2} + \frac{1}{3}} = \\ &= \frac{2 \cdot 3 \cdot \frac{1}{3}}{2 + 3} = \frac{2}{5} = 0.4\end{aligned}$$

Uppgift 14

Stickprov i par. Skillnaderna (Sö-Li) z_i är: 1.2, 0.6, -0.3 , 1.7, 0.9 och de är observationer från $N(\Delta, \sigma)$, där σ är okänd (dvs t -baserat test). Stickprovsmedelvärde och stickprovstandardavvikelse beräknade på de skillnaderna z_i ovan blir: $\bar{z} = 0.82$ och $s_z = 0.746$.

Vi testar:

$$H_0 : \Delta = 0$$

mot

$$H_1 : \Delta > 0.$$

Den observerade testvariabeln (under nollhypotes antagande) blir:

$$t_{obs} = \frac{\bar{z} - 0}{s_z/\sqrt{n}} = \frac{0.82 - 0}{0.746/\sqrt{5}} \approx 2.46,$$

som är en observation från $t(n-1) = t(4)$ -fördelningen, ty temperaturskillnaderna är normalfördelade.

P -värdet är ca 0.035, om man använder sig av miniräknaren.

Om man däremot använder sig av Tabell 3, för $t(4)$ -fördelningen, då ser man att P -värdet för $t_{obs} = 2.46$ ligger mellan 0.025 och 0.05. Dvs., vi har:

$$0.025 < P\text{-värdet} < 0.05.$$

Tolkningen: Eftersom P -värdet < 0.05 så kan man förkasta H_0 på 5 %-nivån. Dvs, man kan påstå att utomhustemperaturen i Södertälje är signifikant högre än på Liljeholmstorget (med chansen på 5 % att detta uttalande kan vara felaktigt, baserad på det observerade datamaterialet). Men på t.ex. 1 %-nivån kan vi dock inte påstå att utomhustemperaturen i Södertälje är signifikant högre än på Liljeholmstorget (dvs med chansen på endast 1 % att detta uttalande kan vara felaktigt), ty P -värdet $> 0.025 > 0.01$.

Uppgift 15

Test av given fördelning. Vi testar följande noll- och mothypoteser:

$$H_0 : p_1 = 0.3, p_2 = 0.5, p_3 = 0.2,$$

mot

$$H_1 : \text{andelen i } H_0 \text{ stämmer ej.}$$

Signifikansnivån är $\alpha = 0.05$.

Den observerade testvariabeln Q_{obs} (under antagandet att nollhypotesen är sann) blir:

$$\begin{aligned}
Q_{obs} &= \frac{(56 - 150 \cdot 0.3)^2}{150 \cdot 0.3} + \frac{(86 - 150 \cdot 0.5)^2}{150 \cdot 0.5} + \frac{(8 - 150 \cdot 0.2)^2}{150 \cdot 0.2} = \\
&= \frac{(56 - 45)^2}{45} + \frac{(86 - 75)^2}{75} + \frac{(8 - 30)^2}{30} = \\
&\simeq 2.69 + 1.61 + 16.13 \simeq 20.4.
\end{aligned}$$

Kritiska värdet som man ska jämföra den observerade testvariabeln Q_{obs} är lika med:

$$\chi^2_{\alpha}(r-1) = \chi^2_{0.05}(2) = 5.99,$$

där $r = 3$ är antal grupper.

Beslutsregeln: förkasta H_0 om $Q_{obs} > \chi^2_{0.05}(2)$.

Slutsatsen: H_0 kan förkastas på nivån 5 %, ty $20.4 > 5.99$.

Tolkningen: på 5 % signifikansnivån kan vi påstå att studenterna i StorStad tränar mer än vad som är vanligt. Deh här slutsatsen följer från en snabb analys av observerade versus förväntade datamaterialet och speciellt stor skillnad betraktar vi i den sista gruppen (dvs., 8 observerade vs 30 förväntade i gruppen *tränar inte alls*). I andra ord är det många fler tränas än det förväntas/än det vad som är vanligt. Resultatet är signifikant på nivån 5 %.

Uppgift 16

Vi har följande resonemang. Om s.v. $\xi \in Po(n)$, där n är ett icke-negativt heltal, så gäller det att

$$P(\xi \leq n) = \sum_{k=0}^n P(\xi = k) = |\xi \in Po(n)| = \sum_{k=0}^n \frac{n^k}{k!} e^{-n} = e^{-n} \sum_{k=0}^n \frac{n^k}{k!}$$

Dvs, det som står i lydelsen är ju alltså som följande:

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \lim_{n \rightarrow \infty} P(\xi \leq n),$$

där $\xi \in Po(n)$, n är ett icke-negativt heltal.

Å andra sidan s.v. $\xi \in Po(n)$ kan representeras som en summa

$$\xi = \xi_1 + \xi_2 + \dots + \xi_n,$$

där $\xi_k \in Po(1)$, $k = 1, 2, \dots, n$, samtliga är oberoende stokastiska variabler. Därmed, nu, om n är tillräckligt stor så har vi enligt CGS (Centrala gränsvärdessatsen) att approximativt gäller följande:

$$\xi = \xi_1 + \xi_2 + \dots + \xi_n \sim N(n\mu, \sigma\sqrt{n}),$$

där $\mu = E(\xi_k) = 1$, $\sigma = \sqrt{V(\xi_k)} = \sqrt{1} = 1$, dvs

$$\xi \sim N(n, \sqrt{n}),$$

approximativt, och approximationen stämmer ju bättre desto större n är. Och, därför, har vi att

$$P(\xi \leq n) \simeq |\text{ty s.v. } \xi \sim N(n, \sqrt{n})| \simeq \frac{1}{2},$$

för alla stora och icke-negative heltal n , och ju närmare $1/2$ desto större n är (enligt CGS). Detta tillsammans med resonemanget ovan bevisar uttrycket i lydelsen.