

TMS051: Matematisk statistik för Z, del B

Tentamen 18 augusti 2004 f V

Tillåtna hjälpmedel är räknedosa utan information om kursen i minnena, Beta samt kursens formel- och tabellsamling.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p.

Jour är Oskar Sandberg, ankn 5336. Examinator är Tommy Norberg, ankn 3528 eller 0730 794209.

Tänk på att ofullständiga uträkningar eller motiveringar till resultat eller beräkningar lätt leder till poängavdrag för den som har ambitionen att uppnå betyget 4:a eller bättre.

Uppgifter

1. Utmed ett gatuavsnitt är tre bilar parkerade. Totalt finns sju parkeringsplatser. Om man antar att de tre bilarnas parkeringsplatser är slumpmässigt valda bland de sju, hur stor är då sannolikheten att de står bredvid varandra? (3 p)
2. Antag att händelserna $A_1, A_2, \dots, A_n$ är en partition av utfallsrummet. Visa att

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

gäller för varje händelse B . (4 p)

3. Tabellen nedan ger sannolikheterna (sånär som på en) för de möjliga utfallen av en två-dimensionell diskret stokastisk variabel (X, Y) , sådan att $0 \leq X \leq 4$ och $0 \leq Y \leq 3$.

		Y			
		0	1	2	3
X	0	.08	.07	.04	.00
	1	.06	.15	.05	.04
	2	.05	?	.10	.06
	3	.00	.01	.05	.06
	4	.00	.01	.05	.06

- (a) Beräkna $P(X = 2, Y = 1)$? (1 p)
- (b) Beräkna $P(X = 2)$ och $P(Y = 1)$. (1 p)
- (c) Avgör, endast m.h.a svaren på de två deluppgifterna ovan, om X och Y kan vara oberoende? (1 p)

4. En viss typ av komponenter har felbenägenheten $r(t) = t^{1.5}$ där $t \geq 0$ är funktions-tiden räknad i år. Bestäm (a) sannolikheten att komponenten fungerar felfritt i 2 år, (b) tätheten i komponentens livslängdsfördelning. (4 p)

5. Täthetsfunktionen för en Paretofördelad stokastisk variabel X , som bara antar värden ≥ 2 , är

$$f(x) = \frac{\alpha}{2} \left(\frac{x}{2}\right)^{-(\alpha+1)}$$

för $x \geq 2$.

- (a) Punktskatta α m.h.a följande 8 oberoende observationer: 2.31, 2.10, 2.29, 2.38, 2.44, 2.54, 2.64, 2.33. (3 p)

- (b) Ungefär hur stor är sannolikheten $p = P(X > 2.50)$. (1 p)

6. I syfte att uppskatta produktionskostnaden av en viss teknisk kemikalie, tillverkades 10 försöksserier om 10 liter. Kostnaden för dessa var 10.50, 11.00, 10.80, 9.70, 9.90, 10.40, 11.20, 10.30, 9.80 och 10.50 USD/liter. Punkt- och intervallskatta den förväntade literkostnaden μ . Önskad konfidensgrad: ca 99%. (4 p)

7. Man har gjort 7 oberoende bestämningar av halten jordbunden kväve i en åker. Därvid erhöles medelvärdet $\bar{x} = 428.1$ och standardavvikelsen $s = 30.31$ i någon (för en statistiker irrelevant) enhet. Åkern betraktas som kväverik om dess medelhalt $\mu > 400$. Medelvärdet verkar ju vara klart större än μ , men är det i statistisk mening signifikant större? (4 p)

8. Man undersökte hur nötningen beror av arbetstemperaturen hos en viss typ av kullager. Följande data uppmättes:

x	200	300	400	500	600
y	3.0	3.9	5.2	6.1	6.8

Här är x temperaturen i Fahrenheit och y är nötningen mätt i mg per 100 arbetstim-mar. Punkt- och intervallskatta nötningen vid temperaturen 400 Fahrenheit. (4 p)

Lycka till!

TMS051: Matematisk statistik för Z, del B

Tentamen 14 april 2004 fm V

Tillåtna hjälpmedel är räknedosa utan information om kursen i minnena, Beta samt kursens formel- och tabellsamling.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p.

Jour och examinator är Tommy Norberg, ankn 3528 eller 0730 794209.

Observera att ofullständiga uträkningar eller motiveringar till resultat eller beräkningar lätt leder till poängavdrag för den som har ambitionen att uppnå betyget 4:a eller bättre.

Uppgifter

1. En standard kortlek består av 52 kort indelade i fyra färger om 13 kort i varje. Varje färg innehåller kort av valören 2–14. Beräkna sannolikheten att en hand bestående av fem slumpvis utvalda kort

(a) bara innehåller kort ur samma färg, (2 p)

(b) innehåller tre kort av en valör och två kort av en annan valör. (2 p)

2. Skriv ned de tre sannolikhetsaxiomen samt bevisa att additionsregeln

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

gäller för två godtyckliga händelser A_1, A_2 . (4 p)

3. En mekaniskt systems livslängd antas följa en s.k Weibull-fördelning med parametrar $\alpha, \beta > 0$. Beräkna systemets felintensitet ("hazard rate"). Ledning: Weibull-fördelningens täthet är

$$f(t) = \alpha \beta t^{\beta-1} e^{-\alpha t^\beta}$$

för $t > 0$. (4 p)

4. Följande "computation formula" gäller för kovariansen:

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

Visa den. Argumentera sedan för påståendet att $\text{Cov}[X, Y] = 0$ om X, Y är oberoende. (3 p)

(Vänd!)

En tillverkare av säkringar påstår, för en viss modell, att om överbelastningen är 20%, så löser säkringen ut efter 12.40 minuter i medel. För att undersöka detta påstående gjordes ett slumpmässigt urval av 20 säkringar ur ett stort parti. Dessa 20 säkringar utsattes för 20% överlast och man mätte tiden tills säkringen bröt. Därvid erhöles medelvärdet 13.88 minuter och standardavvikelsen 2.48 minuter. Resultatet av detta experiment ska analyseras i de tre följande uppgifterna. Analysen ska göras under antagandet att tiden tills dess att en säkring bryter är (åtminstone approximativt) normalfördelad med väntevärdet μ och standardavvikelsen σ .

5. Man är intresserad av den tid $t_{0.99}$, som är sådan att 99% av alla säkringar som är utsatta för 20% överlast bryter innan och 1% efter. Tyvärr går den inte att räkna ut enbart utifrån tillverkarens uppgifter och man vill därför passa på att skatta den. Beräkna en rimlig skattning av $t_{0.99}$. (4 p)
6. (a) Intervallskatta σ uppåt. (2 p)
 (b) Intervallskatta μ symmetriskt. (2 p)
 I båda fallen ska konfidensen vara ca 95%.
7. Antag att du p.g.a tidigare negativa erfarenheter har anledning att misstro tillverkarens uppgift att säkringarna i medel bryter efter 12.40 minuter och att du nu vill utnyttja resultatet av detta experiment till att visa att säkringarna i verkligheten bryter senare. Formulera lämpligt hypotestest och utför detta på nivån 1%. (4 p)
8. Vid ett senare tillfälle gjordes en större studie av tiden tills dess att en säkring bryter vid 20% överlast. Därvid belastades 100 säkringar samtidigt med 20% överlast. Man fann efter 16 minuters testtid att 70 st hade brutit sina resp kretsar. Punkt- och intervallskatta sannolikheten p_{16} att en säkring utsatt för 20% överlast fungerar kortare tid än 16 minuter. (3 p)

Lycka till!

TMS051: Matematisk statistik för Z, del B

Tentamen 13 december 2003 em V

Tillåtna hjälpmedel är räknedosa utan information om kursen i minnena, Beta samt kursens formel- och tabellsamling.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p.

Jour och examinerator är Tommy Norberg, ankn 3528 eller 0730 794209 till kl 16⁴⁵.

Observera att ofullständiga uträkningar eller motiveringar till resultat eller beräkningar ofelbart leder till poängavdrag för den som har ambitionen att uppnå betyget 4:a eller bättre, t.ex gäller att svar av typen ja, nej (j.f.r uppgift 4a) ej godkänns utan korrekt motivering.

Uppgifter

1. Man har utvecklat ett nytt test som, för individer i åldersgruppen 50+, ska detektera en viss typ av arthrititis. Känt är att ca 12% av individerna i åldersgruppen 50+ har besvär av denna typ av arthrititis. Man utsatte individer om vilka man redan visste att de besvärades av denna typ av arthrititis för testet och det visade sig då att testet indikerade arthrititis korrekt i ca 87% av fallen. När man sedan testade garanterat friska individer i åldersgruppen, visade det sig att testet indikerade arthrititis felaktigt i ca 6% av fallen. Bestäm approximativt sannolikheten att en testad individ besvärades av denna typ av arthrititis givet att testet indikerat att så är fallet. (3 p)
2. I snitt fungerar 95% av alla bläckstråleskrivare av ett visst fabrikat och typ direkt efter installation i ett hem. Övriga behöver någon form av intrimning. En försäljare säljer under januari månad 7 st sådana skrivare. (a) Hur stor är sannolikheten att minst 6 av dessa fungerar vid installationen? Antag att försäljaren under ett halvår säljer exakt 7 st sådana skrivare varje månad. (b) Hur stor är sannolikheten att minst 6 av dessa 7 fungerar vid installationen i alla sex månaderna? Sammanlagt säljer alltså försäljaren under halvåret 42 st sådana skrivare. (c) Hur stor är sannolikheten att minst 36 av dessa fungerar vid installationen. (3 p)
3. Visa den s.k "glömskeegenskapen" för exponentialfördelningen, alltså att

$$P(T > s + t | T > s) = P(T > t)$$

gäller för exponentialfördelade stokastiska variabler T . (4 p)

4. (a) Gäller centrala gränsvärdessatsen för Cauchydata, alltså för data som följer tätheten

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{för} \quad -\infty < x < \infty \quad (2 \text{ p})$$

- (b) Låt X, Y vara bivariat normalfördelade med väntevärden $\mu_x = 3.5$, $\mu_y = 7.5$, standardavvikelser $\sigma_x = 1.5$, $\sigma_y = 2$ och korrelation $\rho = 0.75$. Låt $Z = 2X - Y$. Beräkna Z varians. (2 p)

(Vänd!)

5. Under de senaste 10 åren har antalet allvarliga olycksfall i arbetet varit

8 6 12 15 12 4 9 7 20 10

Data gäller för en viss region i södra Sverige. Antag att antalet olyckor av denna typ under ett givet år följer en Poissonfördelning med parameter λ och att antalet olyckor under olika år är oberoende. Härled med någon metod en punktskattning $\hat{\lambda}$ av λ . Är $\hat{\lambda}$ väntevärdesriktig? Beräkna det observerade utfallet av $\hat{\lambda}$. Diskutera modellen. Är det t.ex något i den som du tycker är tvivelaktigt och som möjligen gör resultatet av analysen tveksamt? Inga utsvävningar, tack, utan formulera dig kort och koncist. (4 p)

6. En viss typ av tillverkning har felsannolikheten $p = 0.15$. Man tror att man medelst diverse smarta justeringar kan få ned felsannolikheten till ca 0.10 och man vill "designa" ett försök så att det finns en rimlig chans ($\geq 90\%$) att detektera en förbättring samtidigt som man håller risken (sannolikheten) att göra ett s.k typ-I-fel mindre än 5%. Redogör för hur försöket ska genomföras och för hur analysen av försöksresultatet ska göras samt vilka fel som kan göras och hur stora riskerna är för dessa. (4 p)
7. En hyfsad modell för tiden tills en viss typ av halvledarkomponent upphör att fungera tillfredsställande är $\text{Exp}(\lambda)$. Under en tid har man samlat på sig observationer av $n = 25$ sådana komponenters resp funktionstider. Observerat medelvärde och standardavvikelse är

$$\bar{x} = 10719.0 \quad \text{och} \quad s = 10366.9$$

(tidsenheten är timmar). Intervallskatta parametern λ för den aktuella typen av halvledare. (4 p)

8. I den linjära regressionsmodellen $Y|x \sim N(\beta_0 + \beta_1 x, \sigma)$ gäller att variansen skattas av

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

där $b_1 = S_{xy}/S_{xx}$ och $b_0 = \bar{y} - b_1 \bar{x}$. Visa att

$$(n-2)s^2 = S_{yy} - S_{xy}^2/S_{xx} \quad (4 \text{ p})$$

1. Totalt finns $\binom{7}{3} = 35$ olika sätt att placera ut de tre bilarna på de sju platserna. Av dessa har 5 egenskapen att alla bilarna står bredvid varandra. Således är den sökta sannolikheten $p = 5/35 = 1/7 \approx 0.143$.
2. Att $A_1, \dots, A_n$ partitionerar utfallsrummet Ω betyder att $A_i \cap A_j = \emptyset$ då $i \neq j$ och $\bigcup_i A_i = \Omega$. Vi får

$$P(B) = P\left(B \cap \bigcup_i A_i\right) = P\left(\bigcup_i B \cap A_i\right) = \sum_i P(B \cap A_i) = P(A_i)P(B|A_i)$$

Den 3:e likheten följer ifrån axiomen för sannolikheter och den 4:e är en konsekvens av definitionen av betingning.

3. (a) $P(X = 2, Y = 1) = 0.06$, eftersom summan av alla sannolikheter i tabellen måste bli 1.00. (b) $P(X = 2) = 0.05 + 0.06 + 0.10 + 0.06 = 0.27$ och $P(Y = 1) = 0.07 + 0.15 + 0.06 + 0.01 + 0.01 = 0.30$. (c) X och Y kan ej vara oberoende eftersom $P(X = 2)P(Y = 1) = 0.27 \cdot 0.30 = 0.081 \neq 0.06 = P(X = 2, Y = 1)$.
4. Medelst integration fås överlevnadsfunktionen $R(t) = e^{-\int_0^t r(s) ds} = e^{-0.4 t^{2.5}}$. Således är (a) $P(T > 2) = R(2) = e^{-0.4 \cdot 2^{2.5}} = 0.104$. Fördelningsfunktionen är $F(t) = 1 - R(t) = 1 - e^{-0.4 t^{2.5}}$. Således är (b) tätheten $f(t) = F'(t) = t^{1.5} e^{-0.4 t^{2.5}}$. Den som kollar i formelsamlingen ser att detta är en Weibullfördelning med parametrar $\alpha = 0.4$ och $\beta = 2.5$.
5. (a) Låt $x_1, \dots, x_n$ beteckna observationerna. Trolighetsfunktionen är

$$L(\alpha) = \prod_i f(x_i) = \prod_i \frac{\alpha}{2} \left(\frac{x_i}{2}\right)^{-(\alpha+1)} = \left(\frac{\alpha}{2}\right)^n \left(\prod_i \frac{x_i}{2}\right)^{-(\alpha+1)}$$

och dess logaritm är $\mathcal{L}(\alpha) = n \ln(\alpha/2) - (\alpha + 1) \sum_i \ln(x_i/2)$. Sedvanligt sökande efter maximum ger nu

$$\frac{d\mathcal{L}(\alpha)}{d\alpha} = 0 \Rightarrow \frac{n}{\alpha} - \sum_i \ln(x_i/2) = 0 \Rightarrow \alpha = \frac{n}{\sum_i \ln(x_i/2)}$$

Insättning av observationerna ger nu att trolighetsskattningen av α är $\hat{\alpha} = 5.837$. Ett annat sätt att lösa uppgiften är att beräkna momentskattningen av α . Då behövs

$$E[X] = \int_2^\infty x f(x) dx = \int_2^\infty \alpha 2^\alpha x^{-\alpha} dx = \frac{2\alpha}{\alpha - 1}$$

förutsatt att $\alpha > 1$. Momentskattningen existerar alltså bara om $\alpha > 1$. Den fås genom att lösa ut α ur likheten $\bar{x} = E[X]$. Vi får

$$\bar{x} = \frac{2\hat{\alpha}}{\hat{\alpha} - 1} \Rightarrow \hat{\alpha} = \frac{\bar{x}}{\bar{x} - 2} = \frac{2.37875}{2.37875 - 1} = 6.281$$

(b) Lättast är nog att helt enkelt skatta p med antalet observationer över 2.50 delat med totala antalet observationer. Då fås $\hat{p} = 2/8 = 0.25$. (Denna skattning är väntevärdesriktig.) Den som skattat α kan ju beräkna $P(X > x) = \int_x^\infty f(x) dx = \left(\frac{x}{2}\right)^{-\alpha}$ och sedan skatta p med

$$\hat{p} = 1.25^{-\hat{\alpha}} = 0.272 \text{ resp } 0.246$$

beroende på typ av skattning som gjorts i deluppgift (a).

6. Räkna först ut att $\bar{x} = 10.41$, $s = 0.5043$. I t -tabellen (9 frihetsgrader) avläser vi $P(T > 3.25) = 0.005$. (Obs: $0.005 + 0.005 = 0.01$ och $1 - 0.99 = 0.01$.) Således gäller att $\mu = 10.41 \pm 3.25 \cdot 0.5043/\sqrt{10} = 10.41 \pm 0.52$ USD/liter.
7. Vi ska se om testet av $H_0 : \mu \leq 400$ mot $H_1 : \mu > 400$ förkastar på någon rimligt låg nivå. Vi jämför då teststatistikans utfall $(\bar{x} - \mu_0)/(s/\sqrt{n}) = 28.1/(30.31/\sqrt{7}) = 2.453$ med kvantilerna i t -fördelningen med 6 frihetsgrader. En titt i tabellen ger att utfallets P -värde är aningen mindre än 0.025 ($t_{0.025} = 2.450 < 2.453$). Testet förkastar alltså på nivån 5%, men ej på nivån 1%. Om det ej finns någon anledning att kräva osedvanligt hög säkerhet, så kan man påstå att det uppmätta medelvärde är signifikant större än 400. M.a.o, man kan alltså påstå att åkern är kväverik ($\mu > 400$) och felrisken är då $\leq 5\%$.
8. Slå upp linjär regression i formelsamlingen och följ med. Modellen är $y = \beta_0 + \beta_1 x + \epsilon$, där $\epsilon \sim N(0, \sigma)$. Ur data fås $\bar{x} = 400$, $\bar{y} = 5$, $S_{xx} = \sum_i (x_i - \bar{x})^2 = 100000$, $S_{yy} = \sum_i (y_i - \bar{y})^2 = 9.7$ samt $S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = 980$. Vi kan nu bestämma skattningarna av β_0 och β_1 . De är

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{980}{100000} = 0.0098, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 1.08$$

Vi kan nu beräkna de s.k. residualerna $\hat{\epsilon}_i = y_i - \hat{y}_i$, där $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$:

x	200	300	400	500	600
y	3.00	3.90	5.20	6.10	6.80
$\hat{y}$	3.04	4.02	5.00	5.98	6.96
$\hat{\epsilon}$	-0.04	-0.12	0.20	0.12	-0.16

och fortsätter med skattningen av σ , som fås ur $s^2 = \frac{1}{n-2} \sum_i \hat{\epsilon}_i^2 = 0.032 \approx 0.1789^2$. Punktskattningen av nötningen vid 400 Fahrenheit är $\hat{y} = 5.00$. I formelsamlingen ser vi att denna variabel är $N\left(\beta_0 + \beta_1 x, \sigma \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right)$, där $x = 400$. Vi utnyttjar att $x = \bar{x}$ och konstaterar att konfidensintervallet för y vid 400 Fahrenheit ges av $y = \hat{y} \pm t_{0.025}(n-2)s/\sqrt{n}$. Antalet frihetsgrader i skattningen s^2 av σ^2 är ju $n-2$. I lämplig tabell avläser vi $t_{0.025}(3) = 3.18245$. Således intervallskattningen av y vid 400 Fahrenheit lika med 5.00 ± 0.25 .

1. (a) Antalet sätt att välja färg är 4 och antalet sätt att välja 5 kort ur en färg är $\binom{13}{5}$.

Totala antalet sätt att välja 5 kort är $\binom{52}{5}$. Således blir den sökta sannolikheten enligt multiplikationsprincipen lika med

$$4 \frac{\binom{13}{5}}{\binom{52}{5}} = 4 \frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = \frac{33}{16\,660} \approx 0.001981$$

- (b) Det finns 13 sätt att välja den valör som man ska ha tre utav och därefter 12 sätt att välja den man ska ha två utav. Eftersom det finns fyra kort av varje valör får vi i det tredje valet $\binom{4}{3} \binom{4}{2}$ möjligheter. Den sökta sannolikheten blir

$$13 \cdot 12 \frac{\binom{4}{3} \binom{4}{2}}{\binom{52}{5}} = \frac{6}{4\,165} \approx 0.001441$$

2. Axiomen hittar du i läroboken eller på någon OH-bild från kapitel 2. Additionssatsen kan bevisas på flera sätt. Här följer ett:

$$P(A_1 \cup A_2) = P(A_1 \setminus A_2 \cup A_2) = P(A_1 \setminus A_2) + P(A_2)$$

ty händelserna $A_1 \setminus A_2, A_2$ är ömsesidigt uteslutande. Vidare gäller av en analog orsak att

$$P(A_1) = P(A_1 \setminus A_2 \cup A_1 \cap A_2) = P(A_1 \setminus A_2) + P(A_1 \cap A_2)$$

vilket medför

$$P(A_1 \setminus A_2) = P(A_1) - P(A_1 \cap A_2)$$

Nu följer

$$P(A_1 \cup A_2) = P(A_1) - P(A_1 \cap A_2) + P(A_2)$$

v.s.v.

3. Medelst integration fås tillförlitligheten

$$R(t) = \int_t^\infty \alpha \beta t^{\beta-1} e^{-\alpha t^\beta} dt = e^{-\alpha t^\beta}$$

Felintensiteten är alltså

$$\rho(t) = \frac{f(t)}{R(t)} = \alpha \beta t^{\beta-1}$$

4. Skriv $E[X] = \mu_X$ och $E[Y] = \mu_Y$. Vi får, eftersom E är en linjär operator, att

$$\begin{aligned} \text{Cov}[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] = E[XY - \mu_X Y - X \mu_Y + \mu_X \mu_Y] \\ &= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y \\ &= E[XY] - \mu_X \mu_Y \end{aligned}$$

v.s.v (detta belönas med 2 p). Om X, Y är oberoende, gäller att $E[XY] = E[X] E[Y]$ och då blir kovariansen lika med 0.

5. Både momentmetoden och trolighetsmetoden talar om för oss att följande skattning är klart rimlig:

$$\hat{x}_{0.99} = \bar{x} + z_{0.99} \cdot s = 13.88 + 2.326 \cdot 2.48 = 19.65$$

Vet man inte detta måste man göra sig besväret att härleda antingen momentmetodens skattningar av μ, σ eller trolighetsmetodens. Båda metoderna skattar μ med $\bar{x}$ och σ med $\sqrt{\frac{n-1}{n}} s$. Att byta ut den sistnämnda skattningen mot standardskattningen s av σ är både förståndigt och tidsbesparande.

6. (a) Ur formelsamlingen fås att $(n-1)\frac{s^2}{\sigma^2}$ är χ^2 -fördelad med $n-1$ frihetsgrader. Ur tabell fås att undre 5%-kvantilen är 10.117, och medelst standardomformulering av påståendet

$$P\left((n-1)\frac{s^2}{\sigma^2} > 10.117\right) = 0.95$$

fås intervallskattningen $\sigma^2 < 19 \frac{2.48^2}{10.117} = 11.55 = 3.40^2$. Svaret på deluppgiften är således $\sigma < 3.40$ minuter.

- (b) Ur faktumet att $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ är t -fördelad med $n-1$ frihetsgrader, fås formeln $\mu = \bar{x} \pm t_{0.025}(19) s/\sqrt{n}$ för ett konfidensintervall med den ungefärliga konfidensen 95%. Ur t -tabell fås att kvantilen $t_{0.025}(19) = 2.093$. Insättning ger nu att svaret på deluppgiften är $\mu = 13.88 \pm 1.16$ minuter.

7. Du vill visa att $\mu > 12.40$ minuter, så därför testas du $H_0 : \mu \leq 12.40$ (eller $H_0 : \mu = 12.40$) mot alternativet $H_0 : \mu > 12.40$. Teststatistikan $T = \frac{\bar{x} - 12.40}{s/\sqrt{n}}$ är $t(19)$ -fördelad och har

$$\text{utfallet } t = \frac{13.88 - 12.40}{2.48/\sqrt{20}} = 2.669, \text{ vilket ska jämföras med } 99\text{-kvantilen } t_{0.99}(19) =$$

2.539. Eftersom utfallet är större än kvantilen, förkastas H_0 på nivån 1%. Om du nu i en dialog med tillverkaren av säkringarna påstår att medeltiden tills dess att säkringen löser ut vid 20% överlast är > 12.40 minuter, så är risken att du har fel mindre än ca 1%.

8. Eftersom antalet försök är 100, använder vi standardmetoden som säger att $p_{16} = \hat{p}_{16} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_{16}(1-\hat{p}_{16})}{n}}$. Eftersom inget är sagt om konfidensen väljer jag den till ca 95%. Då blir $\alpha = 5\%$ och $z_{\alpha/2} = z_{0.025} = 1.96$. Insättning av punktskattningen $\hat{p}_{16} = 0.70$ och $n = 100$ ger oss svaret $p_{16} = 0.70 \pm 0.09$. Det finns ingen anledning att ge "felet" med högre noggrannhet.

1. Låt A vara händelsen att den testade individen har arthrits och låt D vara händelsen att testet detekterar arthrits. Vi vet då att

$$P(A) \approx 0.12 \quad P(D|A) \approx 0.87 \quad P(D|A^c) \approx 0.06$$

Bayes formel ger oss den sökta sannolikheten

$$P(A|D) = \frac{P(A)P(D|A)}{P(A)P(D|A) + P(A^c)P(D|A^c)} \approx \frac{0.1044}{0.1044 + 0.0528} \approx 0.66$$

2. (a) $\binom{7}{6} 0.95^6 \cdot 0.05^1 + \binom{7}{7} 0.95^7 \cdot 0.05^0 = 0.2573 + 0.6983 = 0.9556 \approx 0.956$

(b) $0.9556^6 \approx 0.762$

- (c) Den sökta sannolikheten är $p = P(X \geq 36)$, där $X \sim \text{Bin}(42, 0.95) \approx N(39.9, \sqrt{1.995})$. Vi får

$$p \approx 1 - \Phi\left(\frac{35.5 - 39.9}{\sqrt{1.995}}\right) = 1 - \Phi(-3.12) = \Phi(3.12) \approx 0.999$$

med halvtalskorrektur och $\Phi(2.76) \approx 0.997$ utan. En exakt beräkning med Matlab ger resultatet $p = 0.9955$.

3. $P(T > s + t | T > s) = \frac{P(T > s+t, T > s)}{P(T > s)} = \frac{P(T > s+t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t)$

4. (a) Nej, ty Cauchy-fördelningen saknar väntevärde och varians

(b) Vi får

$$\begin{aligned} \text{Var}[Z] &= \text{Var}[2X - Y] = \text{Cov}[2X - Y, 2X - Y] \\ &= 4\text{Cov}[X, X] + \text{Cov}[Y, Y] - 4\text{Cov}[X, Y] \\ &= 4\text{Var}[X] + \text{Var}[Y] - 4\rho\sqrt{\text{Var}[X]\text{Var}[Y]} \\ &= 4 \cdot 1.5^2 + 2^2 - 4 \cdot 0.75 \cdot 1.5 \cdot 2 = 4 \end{aligned}$$

5. Momentmetoden: $\mu = \lambda \Rightarrow \hat{\lambda} = \bar{x}$. Trolighetsmetoden: medelst maximering av trolighetsfunktionen

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\sum_i x_i} / (\prod_i x_i!)$$

fås trolighetsskattningen $\hat{\lambda} = \bar{x}$. Moment- och trolighetsskattningen överensstämmer alltså och är lika med $\hat{\lambda} = \bar{x}$. Att denna skattning är väntevärdesriktig inses av att medelvärdet skattar väntevärdet väntevärdesriktigt. Det observerade utfallet av $\hat{\lambda}$ är $\bar{x} = 103/10 = 10.3$. Poissonmodellen är ofta vettig när det gäller händelser som inträffar relativt sällan, som olyckor förhoppningsvis gör. Det antagande som jag kan tänka mig kanske är tveksamt är oberoendet. Det är ju så att om man under ett år får väldigt många olyckor, så kan det innebära en allmän uppräckning som leder till att antalet olyckor under de följande åren blir lägre än vad de skulle varit annars. Likaså kan ett antal i det närmaste olycksfria år leda till att folks olycksmedvetenhet minskar och att det allmänna slarvet ökar, vilket lätt leder till ett antal olyckor som inte skulle inträffat om vaksamheten hade varit högre. Men det kan ju vara så att olycksförekomsten inte diskuterats offentligt under åren som varit och då är nog oberoendetagandet rimligt.

6. Man ska testa $H_0 : p \geq 0.15$ mot $H_1 : p < 0.15$ i modellen $f \sim \text{Bin}(n, p)$. Praktiskt innebär detta att man ska slumpmässigt välja ut n oberoende enheter tillverkade med den nya metoden för kontroll och beräkna frekvensen felaktiga f . Att sannolikheten för ett typ-I-fel ska vara mindre än 0.05 betyder att testet ska vara på nivån $\alpha = 0.05$. Detta betyder i sin tur att $P(f \leq c) \leq 0.05$ under H_0 , d.v.s då $p \geq 0.15$. Här är c det s.k kritiska värdet och testregeln är förkasta H_0 då $f \leq c$. Hur stort ska man då välja n ? Antag att n är så stort att normalapproximation är tillåten. Då gäller att

$$\begin{aligned} 0.05 &= P(f \leq c) \approx \Phi\left(\frac{c - 0.15n}{\sqrt{0.15 \cdot 0.85n}}\right) \Rightarrow \frac{c - 0.15n}{\sqrt{0.15 \cdot 0.85n}} = -1.645 \\ &\Rightarrow c = 0.15n - 1.645\sqrt{0.15 \cdot 0.85n} = 0.15n - 0.5874\sqrt{n} \end{aligned}$$

Då $p = 0.10$ vill man att sannolikheten att förkasta ska vara iallafall 0.90. Detta ger att

$$0.90 = P(f \leq c) \approx \Phi\left(\frac{c-0.10n}{\sqrt{0.10 \cdot 0.90n}}\right) = \Phi\left(\frac{0.05n-0.5874\sqrt{n}}{0.3\sqrt{n}}\right)$$

Således gäller

$$1.282 = \frac{0.05n-0.5874\sqrt{n}}{0.3\sqrt{n}} \Rightarrow \sqrt{n} = 19.44 \Rightarrow n \approx 378$$

(Det går alltså bra att normalapproximera i detta problemet.) Väljes stickprovsstorleken $n = 380$ och regeln förkasta $H_0 : p \geq 0.15$ till förmån för $H_1 : p < 0.15$ då frekvensen felaktiga $f \leq 45$, så blir nivån $\alpha \approx 0.05$ och upptäcktssannolikheten $1 - \beta \approx 0.90$, vilket var vad som krävdes. Väljes ett större värde på n gäller att förkastelseregeln ska vara $f \leq 0.15n - 0.5874\sqrt{n}$. Då håller testet nivån $\alpha \approx 0.05$ och styrkan $1 - \beta$ är $\approx \Phi\left(\frac{0.05n-0.5874\sqrt{n}}{0.3\sqrt{n}}\right) \gtrapprox 0.90$.

7. Ur formelsamlingen (se Ett exponentialfördelat stickprov) fås att $2\lambda n\bar{x} \sim \chi^2(2n)$. Antalet frihetsgrader är alltså $2n = 50$. Ur tabell $\chi^2(50)$ -fördelningen, ser vi att 0.025- och 0.975-kvantilerna är 32.3574 och 71.4204 (jag väljer konfidensen 0.95 eftersom inget är sagt om denna i problemet). Således gäller

$$P(32.3574 \leq 2\lambda n\bar{x} \leq 71.4204) = 0.95$$

vilket räknas om till konfidensintervallet $0.00006037 \leq 10^6 \lambda \leq 0.00013326$, d.v.s $60.37 \leq 10^6 \lambda \leq 133.26$. Alternativt kan man, eftersom normalapproximationen av medelvärdet av $n = 25$ oberoende exponentialvariabler är ganska bra (j.f.r projektuppgift 1), bilda

$$1/\lambda = \mu = 10719.0 \pm 2.064 \cdot 10366.9/\sqrt{25} = 10719.0 \pm 2073.38$$

(obs $2.064 = t_{0.025}(24)$) och medelst invertering av alla leden komma fram till intervallet $0.00007817 \leq \lambda \leq 0.00011567$ eller $78.17 \leq \lambda \leq 115.67$. Obs att konfidensen i detta intervall endast är approximativ. Att det blev kortare än det exakt beräknade, beror nog mest på att utfallet av s råkade bli mindre än utfallet av $\bar{x}$. Normalt ska vi ha $s \approx \bar{x}$ då data är exponentialfördelade.

8. Vi får

$$\begin{aligned} (n-2)s^2 &= \sum_i (y_i - b_0 - b_1 x_i)^2 \\ &= \sum_i (y_i - \bar{y} + b_1 \bar{x} - b_1 x_i)^2 = \sum_i [(y_i - \bar{y}) - b_1(x_i - \bar{x})]^2 \\ &= \sum_i (y_i - \bar{y})^2 - 2b_1 \sum_i (x_i - \bar{x})(y_i - \bar{y}) + b_1^2 \sum_i (x_i - \bar{x})^2 \\ &= S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx} = S_{yy} - 2S_{xy}^2/S_{xx} + S_{xy}^2/S_{xx} \\ &= S_{yy} - S_{xy}^2/S_{xx} \end{aligned}$$

vilket skulle visas.