MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 26 augusti 2008 fm M

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

Examinator är Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Assistent är Sofia Tapani, ankn 5336.

Sofia (eller Tommy) går att nås per telefon under tentamen.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan. Rättningsprotokoll anslås ej.

Svar och lösningar skall motiveras om ej annat sägs i uppgiften. Omotiverade svar betraktas som gissningar och ger sällan poäng.

Uppgifter

1. Om händelserna A och B vet man att $P(A) = 0.3$ samt att $P(B) = 0.5$.
   (a) Beräkna $P(A \cap B)$ om A och B vore oförenliga. (1 p)
   (b) Beräkna $P(A \cap B)$ om A och B vore oberoende. (1 p)
   (c) Hur stor kan $P(A \cap B)$ maximalt vara? (1 p)

2. I långväga digitala dataöverföringar (t.ex från en satellit som just landat på planeten Mars) är risken för överföringsfel relativt hög och man behöver vidta åtgärder för att hantera detta. Vi ska här studera överföringen av en bit (en bit är den minsta informationsenheten och är antingen en nolla eller en etta). Antag i en specifik applikation att felrisken är ca 2.5%. Antag även att man sänder ungefär 75% ettor och 25% nollor. Givet att en etta tagits emot, hur stor är sannolikheten att det var en etta som sändes? (4 p)

3. Visa att funktionen
$$m(t) = \frac{pe^t}{1 - qe^t} \quad \text{för} \quad t < -\ln q$$
där $q = 1 - p$, är m.g.f för den geometriska fördelningen med parameter $p \in (0, 1)$. Visa sedan att fördelningens väntevärde är $\mu = 1/p$ och att dess varians är $\sigma^2 = q/p^2$. (2+1+1 p)

4. Visa att om du gör n oberoende upprepningar av ett försök i vilket en händelse A antingen inträffar eller icke, och låter x vara antalet gånger A inträffar, så har x massfunktionen
$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{för} \quad x = 0, 1, \ldots, n$$
där p är händelsen A:s sannolikhet. (4 p)

5. Redogör för hur man m.h.a två slumptal u_1, u_2 kan simulera två oberoende N(0,1)-variabler z_1, z_2. Bevis eller motivering behövs ej. (4 p)

6. I en partisympatiundersökning tillfrågades 1296 potentiella väljare. Av dessa föredrog 324 st partiet A och man drog slutsatsen att A skulle få ca 25% av rösterna i det kommande valet. Sedan tillfrågades du, som nyss läst grundkursen i matematisk statistik, hur säker denna slutsats är. Bestäm därför ett nedåt begränsat konfidensintervall för A:s väljarandel med ca 95% konfidensgrad. Jag har med avsikt underlåtit att nämna ett viktigt krav på undersökningen som måste vara uppfyllt. Annars kan man varken hävda att ca 25% av väljarna kommer att rösta på A eller beräkna konfidensgränser. Vilket är detta krav? (2+1 p)

7. För datamängden

$$0.08 \quad 0.57 \quad 0.60 \quad -0.02 \quad 1.68 \quad -0.31 \quad 1.64$$

($\sum x = 4.24$ och $\sum x^2 = 6.2998$, $n = 7$). testa nollhypotesen $H_0 : \mu \leq 0$ mot alternativet $H_1 : \mu > 0$ på nivån 5%. Antag oberoende och likafördelade normalfördelade observationer. (4 p)

8. Försöket som ledde fram till data i ovanstående uppgift upprepades efter ett halvår. Därvid erhölls

$$1.50 \quad -0.22 \quad 1.83 \quad 0.59 \quad 0.95 \quad 1.08 \quad 0.10$$

($\bar{x} = 0.833$ och $s = 0.7324$, $n = 7$). Dock hade man anledning att misstänka att förhållandena runt försöket hade ändrats så att medelvärdet μ nu var större. Styrker de två datamängderna denna misstanke? Antag att de två försöken är oberoende samt att observationernas varians är densamma i bägge försöken. (4 p)

Lycka till!

Svar till vissa uppgifter, MVE090 den 26/8-08

1. (a) 0 (b) 0.15 (c) 0.3
2. 0.99
3. Se läroboken
4. Se läroboken
5. Se simuleringshäftet
6. Kravet är att man gjort ett slumpmässigt urval ur hela väljarpopulationen. Gränsen i konfidensintervallet är 23.0%
7. H_0 kan förkastas på nivån 5%
8. Nej, testet av $H_0 : \mu_2 = \mu_1$ vs $H_1 : \mu_2 > \mu_1$ förkastar inte ens på nivån 10%

MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 17 januari 2008 em M

Obs att detta är även omtentamen för TMS051, del B.

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

Examinator är Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Assistent är Sofia Tapani, ankn 5336.

Jour: Sofia (eller Tommy) går att nås per telefon under tentamen.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan. Rättningsprotokoll anslås ej.

Svar och lösningar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. Betrakta ett slumpmässigt försök i vilket händelserna A, B, C är ömsesidigt uteslutande och sådana att exakt en av dem måste inträffa. T.ex ett tärningskast och $A = \{1, 4\}$, $B = \{2\}$ och $C = \{3, 5, 6\}$. Låt F vara en annan händelse. Visa först allmänt att

$$P(F) = P(A)P(F|A) + P(B)P(F|B) + P(C)P(F|C) \tag{1}$$

och verifiera sedan (1) numeriskt för tärningskastfallet med F som händelsen att ett udda antal prickar erhålls. Antag därvid att tärningen är symmetrisk. Vad är $P(A|F)$? Ange först en formel för beräkning av $P(A|F)$. Räkna sedan ut (med hjälp av formeln) denna betingade sannolikhet i tärningskastexemplet. (5 p)

2. I en studie av risken för utbrott av vattenburen smitta (d.v.s smitta spriden genom dricksvattnet) på Europanivå redogjordes för 86 kända utbrott under åren 1990–2005. Antag en lämplig fördelning för antalet utbrott under ett år, och skatta sannolikheten för minst 2 utbrott under år 2008. (3 p)

3. Härled en formel för simulering av observationer ifrån Paretofördelningen. (3 p)

4. I denna uppgift ska vi studera en två-dimensionell stokastisk variabel N, X, där N är diskret och X är kontinuerlig. Täthet för N, X är funktionen

$$f(n,x) = e^{-\lambda}\frac{\lambda^n}{n!}\frac{1}{\sqrt{2\pi n}\sigma}e^{-\frac{1}{2}(x-n\mu)^2/(n\sigma^2)} \quad \text{för} \quad n = 0, 1, \ldots, \ -\infty < x < \infty$$

där $\lambda > 0$, $-\infty < \mu < \infty$ och $\sigma > 0$. Beräkna den betingade tätheten för X givet $N = n$. Motivera extra noga stegen i beräkningen. Beskriv motsvarande fördelning i ord. (4 p)

5. I en studie av ett nytt mätinstrument mätta man mot en s.k normal med det kända värdet 1. Därvid erhölls i $n = 25$ mätningar $\sum x = 27.71$ och $\sum x^2 = 35.984$. Bestäm ett uppåt begränsat 95% konfidensintervall för mätinstrumentets standardavvikelse. Antag normalfördelade observationer. (3 p)

6. I en simulering av en produktionsanläggning hann man under en dag göra 400 oberoende observationer av tiden det tar för en operatör att plocka ut ett bearbetat råämne ur en maskin och sätta in ett nytt, samt starta maskinen. Därvid erhölls $\sum x = 11\,028.6$ och $\sum x^2 = 314\,610.46$. Beräkna ett 99% konfidensintervall för förväntad operatörstid. Tidsenheten är sekunder, så det kan vara lämpligt att avrunda till en decimal. Ange förutsättningarna för dina beräkningar. (3 p)

7. Tillverkare av allehanda produkter är (bör i alla fall vara) intresserade av sina produkters (förväntade) livstider. Statistik över livstider kan vara både dyr och tidskrävande att få fram. I ett fall studerades 5 enheter tills alla slutat fungera. Man erhöll

   3.48 2.41 1.43 2.75 6.59

   (tidsenhet: år). Man utgår ifrån att data är Weibullfördelade med parametrar α och β. I en verklig arbetssituation skulle din uppgift kanske vara att ML-skatta α och β. Men för att denna uppgift inte ska bli för komplicerad rent numeriskt vill jag att du endast ML-skattar α under antagandet att $\beta = 2$. Skatta sedan förväntad livslängd μ, samt det som i en del sammanhang kallas L20, nämligen den tid som i genomsnitt 80% av alla producerade enheter överlever. Klarar du inte att ML-skatta α, så går det ändå att skatta μ och L20. Gör i så fall det. (5 p)

8. Vid en jämförelse av två algoritmer för lösning av det s.k handelsresandeproblemet upprepade man en beräkning med algoritm A 5 gånger och med algoritm B 7 gånger. Men erhöll olika värden varje gång eftersom algoritmerna innehåller slumpmoment som ger varierande beräkningstider på samma problem. För algoritm A erhölls

   4.53 4.26 5.18 5.24 5.16

   och för algoritm B erhölls

   4.28 4.92 4.81 5.06 4.78 4.79 4.60

   Tidsenhet: minuter. Kan du m.h.a statistiska metoder du lärt i denna kurs påvisa att algoritmerna är olika i något avseende? Antag oberoende normalfördelade observationer. (4 p)

Kortfattade svar eller lösningar till MVE090 den 17/1-08

1. $P(F) = P(F\cap A)+P(F\cap B)+P(F\cap C) = P(A)P(F|A)+P(B)P(F|B)+P(C)P(F|C)$ (den 1a likheten följer ifrån additivitet och att utfallsrummet är $A \cup B \cup C$, den 2a ifrån definitionen av betingad sannolikhet). Formeln är $P(A|F) = P(A)P(F|A)/P(F)$. De numeriska bitarna överlåtes.

2. Poissonfördelning med intensiteten $\lambda \approx 86/16 = 5.375$. Sannolikheten för minst 2 utbrott under år 2008 är $\sum_{k=2}^{\infty} e^{-\lambda}\lambda^k/k! = 1-(e^{-\lambda}+e^{-\lambda}\lambda) \approx 0.970$

3. Enl känd sats ska man lösa x ur $u = F(x)$ eller $u = 1 - F(x)$, där u är slumptalet. För Paretofördelningen gäller att $1 - F(x) = (x_T/x)^\alpha$ för $x > x_T > 0$, så vi får direkt ur den andra ekvationen att $x = x_T/u^{1/\alpha}$.

4. Det gäller här att tänka ut uppdelningen $f(x, n) = p(n)f(x|n)$ och känna igen Poi(λ)-tätheten

$$p(n) = e^{-\lambda}\frac{\lambda^n}{n!}$$

och inse att $N \sim \text{Poi}(\lambda)$. Den betingade tätheten för X givet $N = n$ är således

$$f(x|n) = \frac{1}{\sqrt{2\pi n}\sigma} e^{-\frac{1}{2}(x-n\mu)^2/(n\sigma^2)}$$

Detta är en tätheten i $\text{N}(n\mu, \sqrt{n}\sigma)$-fördelningen. Man kan tänka sig att X är en summa av N oberoende $\text{N}(\mu, \sigma)$-variabler.

5. $s^2 = (35.985 - 27.71^2/25)/24 = 0.2196 = 0.4687^2$. Ur $(n-1)s^2/\sigma^2 \sim \chi^2(n-1)$ konstaterar vi att $P((n-1)s^2/\sigma^2 \geq 13.8484) = 0.95$ (se lämplig tabell), så $\sigma^2 \leq (n-1)s^2/13.8484 = 0.3806 = 0.617^2$ eller $\sigma \leq 0.617$ gäller med konfidensen 95%. Använder man den förenklade metoden i Beta, så ska man få $\sigma < 1.316s = 0.617$

6. Har man 400 observationer vore det mycket ovanligt om inte centrala gränsvärdessatsen kan tillämpas. Den samt faktumet $\sigma \approx s$ ger att approximativt gäller $(\bar{x}-\mu)/(s/\sqrt{n}) \sim \text{N}(0, 1)$. Så med ca 99% konfidens gäller $\mu = \bar{x} \pm 2.576s/\sqrt{n}$. Återstår att beräkna $\bar{x} = 11028.6/400 = 27.572$ och $s^2 = (314610.46 - 11028.6^2/400)/399 = 26.4045 = 5.1385^2$ och, därför, $\mu = 27.6 \pm 0.7$

7. Weibulltätheten är $f(x) = \alpha\beta x^{\beta-1}e^{-\alpha x^\beta}$, så troligheten är

$$L(\alpha, \beta) = \prod_i \alpha\beta x_i^{\beta-1}e^{-\alpha x_i^\beta} = \alpha^n\beta^n \left(\prod_i x_i\right)^{\beta-1} e^{-\alpha\sum_i x_i^\beta}$$

Den logaritmeras och maximeras m.a.p α för fixt $\beta = 2$, enl

$$\log L(\alpha, \beta) = n\log\alpha + n\log\beta + (\beta-1)\sum_i x_i - \alpha\sum_i x_i^\beta$$
$$0 = \frac{\partial \log L(\alpha,\beta)}{\partial\alpha} = \frac{n}{\alpha} - \sum_i x_i^\beta \quad \Rightarrow \quad \alpha = \frac{n}{\sum_i x_i^\beta}$$

ML-skattningen av α är således $\widehat{\alpha} = \frac{n}{\sum_i x_i^\beta} = \frac{5}{70.9540} = 0.0705$, ty $\beta = 2$. I formelsamlingen ses att $\mu = \alpha^{-1/\beta}\Gamma(1 + 1/\beta) = \alpha^{-1/2}\Gamma(1.5)$. I den ser vi också att $\Gamma(1.5) = 0.5\Gamma(0.5) = 0.5\sqrt{\pi}$, så vi får till sist skattningen $\widehat{\mu} = 0.0705^{-1/2}0.5\sqrt{\pi} = 3.338$. Den som inte kan ML-skatta α, beräknar istället $\widehat{\mu} = \bar{x} = 3.332$ och sedan $\widehat{\alpha}$ ur sambandet $\mu = \alpha^{-1/2}\Gamma(1.5)$. Han/hon bör få $\widehat{\alpha} = 0.0707$. L20 är estimatet av 0.20-kvantilen $x_{0.2}$. Den fås teoretiskt genom att lösa $F(x_{0.2}) = 0.20$. Ur $F(x_{0.2}) = 1 - e^{-\alpha x_{0.2}^\beta}$ fås att $x_{0.2} = ((-\log 0.8)/\alpha)^{1/\beta}$, så $\text{L20} = ((-\log 0.8)/0.0705)^{1/2} = 1.779 \approx 1.78$. En sista titt på data ger att resultatet är rimligt (en observation är ju mindre och de övriga fyra större).

8. Man börjar med att notera (räkna ut) att $n_A = 5$, $\bar{x}_A = 4.874$, $s_A = 0.449$ och $n_B = 7$, $\bar{x}_B = 4.749$, $s_B = 0.250$. Använd därvid gärna formlerna $\bar{x} = \left(\sum x\right)/n$ och $s^2 = \left(\sum x^2 - \left(\sum x\right)^2/n\right)/(n-1)$. Ur den hemska formlen på sida 348 får vi att antalet frihetsgrader i jämförelsen ska vara 5. Räkna sedan ut

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_A^2/n_a + s_B^2/n_B}} = 0.506$$

och jämför med $t_{0.10}(5) = 1.476$. Så inte ens med 20% felrisk kan vi påstå att $\mu_a \neq \mu_B$.

MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 24 oktober 2007 em M

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

Examinator är Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Jour är Sofia Tapani, ankn 5336.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan. Rättningsprotokoll anslås ej.

Svar och lösningar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. Du drar två kort från en väl blandad kortlek.

   (a) Visa att sannolikheten att det andra kortet är ett ess är 1/13 om du inte har sett det första.

   (b) Om du däremot har sett att det första kortet är ett ess, vad är då sannolikheten att det andra kortet är ett ess?

   (c) Hur stor är sannolikheten att exakt ett av korten är ett ess?

   (d) Givet att det andra kortet är ett ess, hur stor är sannolikheten att det första också är ett ess?

   Hjälp för den som behöver: En kortlek består av 52 kort. Av dessa är 4 ess. (3 p)

2. Antalet patogena (hälsofarliga) bakterier i Göta Älv antas vara Poissonfördelat med parameter $\lambda = 25$ st per m^3 vatten. I ditt mikroskop betraktar du en droppe innehållande 1 ml älvvatten. Beräkna sannolikheten att din droppe innehåller minst en patogen bakterie. Hjälp för den som behöver: 1 m^3 = 1000 l. (3 p)

3. Härled ett snyggt slutet analytiskt uttryck för mgf för Poi(λ). Visa m.h.a detta att $\mu = \sigma^2 = \lambda$. (4 p)

4. Anta att paret X, Y har en bivariat (tvådimensionell) fördelning med täthet

   $$f(x, y) = c \quad \text{för} \quad 0 < x \leq y < 1$$

   där $c > 0$. Bestäm, för $0 < x < 1$, den betingade tätheten för Y givet $X = x$. (4 p)

5. Den s.k Paretofördelningen, Pareto(u, α), definieras av att fördelningsfunktionen är

   $$F(x) = 1 - \left(\frac{u}{x}\right)^\alpha \quad \text{för} \quad x > u$$

   där $u > 0$ är en känd nivå och $\alpha > 0$. Härled

   (a) momentskattningen (2 p)

   (b) trolighetsskattningen (2 p)

   av parametern α. Härvid förutsätts att du har givet data $x_1, \ldots, x_n$, som kan anses vara oberoende observationer av Pareto(u, α).

6. I en undersökning av partisympatier erhöll partierna A och B frekvenserna $f_A = 476$ resp $f_B = 119$. Totalt tillfrågades 1372 slumpmässigt utvalda väljare och av dessa valde $n = 1359$ att delta i undersökningen (d.v.s svara på frågan ”Vilket parti skulle du rösta på om det var val idag?”). Punkt- och intervallskatta de sanna proportionerna p_A resp p_B som skulle välja att rösta på A och B om det vore val idag (eller snarare den dag de tillfrågades). Varför är den statistiska felmarginalen mindre för parti B än för parti A? Skriv något klokt om bortfallet som bestod av 13 personer. (5 p)

7. I en mätning av tiden det tar för en robot att utföra ett visst produktionsmoment erhölls i $n = 25$ mätningar medelvärdet $\bar{x} = 47.3$ sekunder och variansen $s^2 = 22.84$. Det anses viktigt att variationen i utförandetiderna är liten. Beräkna därför ett 99% uppåt begränsat konfidensintervall för standardavvikelsen σ. Redovisa vilka antaganden du behöver göra för att kunna lösa uppgiften. (4 p)

8. Beräkna ett 95% konfidensintervall för robotens förväntade utförandetid μ. Redovisa vilka antaganden du behöver göra för att kunna lösa uppgiften. (3 p)

Lycka till!

Kortfattade svar eller lösningar till MVE090 den 24/10-07

1. (a) $\frac{4}{52}\frac{3}{51}+\frac{48}{52}\frac{4}{51}=\frac{1}{13}$ (lagen om total sannolikhet)

   (b) $\frac{3}{51}=\frac{1}{17}$

   (c) $\frac{4}{52}\frac{48}{51}+\frac{48}{52}\frac{4}{51}=\frac{32}{221}$ (lagen om total sannolikhet) eller $\frac{\binom{4}{1}\binom{48}{1}}{\binom{52}{2}}=\frac{32}{221}$

   (d) $\frac{\frac{4}{52}\frac{3}{51}}{\frac{4}{52}\frac{3}{51}+\frac{48}{52}\frac{4}{51}}=\frac{1}{17}$ (Bayes formel)

2. Antalet i en droppe är Poi(λ) med $\lambda=25\cdot 10^{-6}$, så sannolikheten för minst en bakterie är $1-e^{-\lambda}\approx\lambda=0.000\,002\,5$ (då λ är mycket liten gäller ju med god noggranhet att $e^{-\lambda}=1-\lambda$)

3. M.g.f för Poisson är $m(t)=\sum_{k=0}^{\infty}e^{kt}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!}=e^{-\lambda}e^{\lambda e^t}=e^{\lambda(e^t-1)}$
   Derivering: $m'(t)=\lambda e^t e^{\lambda(e^t-1)}=\lambda e^t m(t)$ och $m''(t)=\lambda e^t m(t)+\lambda e^t m'(t)$
   Så, $m'(0)=\lambda$ och $m''(0)=\lambda+\lambda^2$ och det följer att $\mu=\lambda$ samt $\sigma^2=\lambda+\lambda^2-\lambda^2=\lambda$

4. Givet $x\in(0,1)$ är inget y-värde mer troligt än något annat. Mao, $Y|x$ är likformigt fördelad, så vi måste ha att $f_{Y|x}(y)=\frac{1}{1-x}$ för $x\le y<1$
   Alternativ: $f_X(x)=\int_x^1 c\,dy=c(1-x)$ och $f_{Y|x}(y)=\frac{c}{c(1-x)}=\frac{1}{1-x}$ för $x\le y<1$

5. (a) Obs att $f(x)=F'(x)=\frac{\alpha}{u}\left(\frac{u}{x}\right)^{\alpha+1}$. I formelsamlingen ses att $\mu=\frac{\alpha}{\alpha-1}u$ så $\alpha=\frac{\mu}{\mu-u}$. Härur fås att momentskattningen är $\widehat{\alpha}=\frac{\bar{x}}{\bar{x}-u}$. En förutsättning för att detta ska fungera är att $\alpha>1$. Annars existerar ju inte väntevärdet.

   (b) Trolighetsfunktionen är $L(\alpha)=\prod_i\frac{\alpha}{u}\left(\frac{u}{x_i}\right)^{\alpha+1}=\left(\frac{\alpha}{u}\right)^n\left(\prod_i\frac{u}{x_i}\right)^{\alpha+1}$
   Logaritmera: $\mathcal{L}(\alpha)=n\ln\alpha-n\ln u+(\alpha+1)(n\ln u-\sum_i\ln x_i)$
   Derivera: $\mathcal{L}'(\alpha)=\frac{n}{\alpha}+n\ln u-\sum_i\ln x_i$
   Detta $=0$ precis då $\frac{n}{\alpha}=\sum_i\ln x_i-n\ln u$, d.v.s då $\alpha=\frac{n}{\sum_i\ln x_i-n\ln u}=\frac{n}{\sum_i\ln(x_i/u)}$
   Trolighetsskattningen är alltså $\widehat{\alpha}=\frac{n}{\sum_i\ln(x_i/u)}=\frac{1}{\frac{1}{n}\sum_i\ln(x_i/u)}$

   (Man ser att dessa skattningar ger olika resultat. Trolighetsskattningen rekommenderas. Den har bättre egenskaper.)

6. Bortfallet är litet. Om samtliga, t.ex, skulle svarat A, så skulle p_A skattats med $\frac{476+13}{1372}=0.356$ och p_B med $\frac{119-13}{1372}=0.0773$, istället för som nu $\hat{p}_A=\frac{476}{1359}=0.350$ resp $\hat{p}_B=\frac{119}{1359}=0.0876$, men att enbart parti A-sympatisörer avstått från att deltaga är knappast troligt. Antar man att bortfallet har ungefär samma sympatier som de svarande, så blir skattningarna ungefär detsamma som de skulle blivit om alla tillfrågade deltagit i undersökningen.
   Inget är sagt om konfidensgraden i uppgiften, så vi väljer att intervallskatta med säkerheten ca 95%. Då är $\alpha=0.05$ och $z_{\alpha/2}=z_{0.025}=1.96$ samt $p_A=0.350\pm 1.96\sqrt{\frac{0.350\cdot 0.650}{1359}}=0.350\pm 0.025$ och $p_B=0.0876\pm 1.96\sqrt{\frac{0.0876\cdot 0.9124}{1359}}=0.0876\pm 0.0150$.
   Att den statistiska felmarginalen är större för parti A än för parti B beror på att funktionen $p\curvearrowright p(1-p)$ växer från 0 till 0.25 då p växer från 0 till 0.5 och sedan avtar ned mot 0. M.a.o $p_B(1-p_B)<p_A(1-p_A)$. Vi vet ju att $p_B\approx 0.09<0.35\approx p_A$.

7. Det är viktigt att data är approximativt normalfördelade. Centrala gränsvärdessatsen kan man inte luta sig så väldigt starkt emot när det är σ eller σ^2 man ska intervallskatta. (J.f.r läroboken, sida 281, nedersta stycket.) I formelsamlingen ser vi att $\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$, så sannolikheten att händelsen $\frac{(n-1)s^2}{\sigma^2}\ge\chi^2_{0.99}(24)=10.8564$ inträffat är 0.99. Vi vänder på olikheten och får $\sigma^2\le\frac{(n-1)s^2}{10.8564}=50.4919\Rightarrow\sigma\le 7.11$. Notera att σ skattas av $s=\sqrt{22.84}=4.78$.

8. Här går det bättre att hänvisa till centrala gränsvärdessatsen och räkna som om data vore normalfördelade, trots att vi inte alls vet något om datas fördelning. Men obs att det finns besvärliga fördelningar som man behöver väldigt stora stickprov på för att

centrala gränsvärdessatsen ska fungera bra. Så vi kan inte vara bergsäkra på att $\mu = \bar{x} \pm t_{\alpha/2}(n-1)s/\sqrt{n} = 47.3 \pm 2.0639 \cdot \sqrt{22.84/25} = 47.3 \pm 1.97$ (eller $45.33 \leq \mu \leq 49.27$) har den önskade säkerheten ca 95%.