MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 14 januari 2010 em

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.

Övningsledare är Anna Rudvik, ankn 5338 eller 0730 57 96 26.

Jour: Anna och/eller Tommy är tillgänglig per telefon.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a.

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

Uppgifter

1 I en kursenkät deltog 56 elever. Frågorna 3 och 7 var s.k valfrågor där de svarande skulle välja ett av alternativen a, b och c. Kalla en elev positiv P (resp negativ N) om denna valt a eller b (resp b eller c) på båda frågorna. Antalet negativa resp positiva elever var 30 resp 35, och 50 elever var negativa eller positiva.

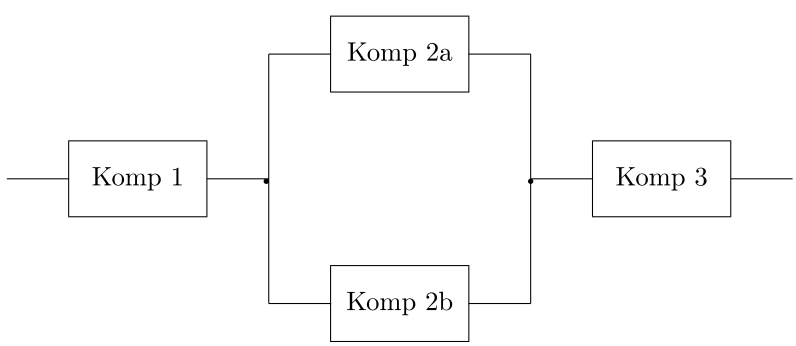
(a) Hur många elever var både negativa och positiva? (1 p)

(b) Hur många elever var bara positiva? (1 p)

2 Man undersökte en stor populations köpvanor under en kalendermånad. Ca 20% besökte IKEA och ca 15% av dessa besökte även Teknikmagasinet någon gång under perioden. Ca 48% besökte inte någon av köpinrättningarna. Av de som någon gång under perioden besökte Teknikmagasinet, hur stor andel besökte även IKEA? (4 p)

3 Antalet olyckor med dödlig utgång utefter en relativt belastad sträcka i det nationella vägnätet var under åren 2004–2008, $1 + 2 + 3 + 0 + 1 = 7$. Ange en lämplig modell, skatta dess parameter väntevärdesriktigt och beräkna m.h.a modellen en skattning av sannolikheten för minst en olycka med dödlig utgång under år 2009. Bedöm rimligheten av ditt resultat genom att jämföra med en annan lätt beräknad skattning av samma sannolikhet. (4 p)

4 Fyra komponenter är ihopkopplade i tillförlitlighetsavseende enl följande schema.



Efter 10 000 timmars drift bedöms tillförlitligheten vara 0.99 för komponenterna 1 och 3, samt 0.90 för komponenterna 2a och 2b. Beräkna den resulterande tillförlitligheten för hela tre-komponentsystemet. Obs Ge svaret med 2 decimalers noggranhet. (3 p)

5 En detalj i en konstruktion har felbenägenheten $z(x) = cx^{0.1}$, för något $c > 0$, där $x > 0$ betecknar gångtid i km. Man har data $x_1, \ldots, x_n$ på $n = 5$ oberoende gångtider till fel:

$$x_1 = 1129241 \quad x_2 = 9490772 \quad x_3 = 4919829 \quad x_4 = 2726391 \quad x_5 = 7805152$$

Härled ett uttryck för trolighetsskattningen $\hat{c}$ av c som funktion av data $x_1, \ldots, x_n$. Sätt in data ovan, och beräkna $\hat{c}$ (hjälp: $\sum_i x_i^{1.1} = 1.25 \cdot 10^8$). (Obs att det p.g.a hjälpen är viktigt att trolighetsskattningen är korrekt härledd. Var därför noga med detaljerna och motiveringarna av dessa.) (4 p)

6 IQ-test utformas så att fördelningen för resultatet i en stor population (som hela Sverige) av individer med ungefär samma ålder ska bli hyfsat lik en normalfördelning med väntevärde $\mu = 100$ och standardavvikelse $\sigma = 15$. Man valde slumpmässigt ut $n = 15$ individer med en viss egenskap och utsatte dessa för ett IQ-test. Syftet var att försöka få bekräftat att individer med denna egenskap har en tendens att ha högre IQ än genomsnittet. Möjligen kan det även vara så att standardavvikelsen är lägre i denna subpopulation. Man erhöll medelvärdet $\bar{x} = 105.9$ och standardavvikelsen $s = 10.55$.

(a) Verifierar undersökningen på 5%- eller 1%-nivån hypotesen att standardavvikelsen är lägre i denna subpopulation? (3 p)

(b) Verifierar undersökningen hypotesen att medelvärdet i denna subpopulation är större än medelvärdet i hela populationen? Formulera korrekt ett statistiskt test, beräkna P-värde så gott det låter sig göras m.h.a formelsamlingen och förklara resultatet så att en novis på området förstår. (4 p)

7 I en nyligen genomförd partisympatiundersökning skattades konstellationen X:s väljarandel till 52.6%. Det var 1516 röstberättigade som besvarade enkäten. Vi antar att dessa var slumpmässigt utvalda på så sätt att alla potentiella väljare hade samma chans att komma med (vilket inte alltid är sannt) och att skattningen av väljarandelen var den naiva $\hat{p}_X = f_X/n$, där f_X betecknar frekvensen sympatisörer och n antalet tillfrågade (inte heller detta är nödvändigtvis sannt för det finns intelligentare sätt att skatta andelar). Beräkna ett 95% konfidensintervall för den sanna andelen väljare som sympatiserade med konstellationen X vid tillfället ifråga. (3 p)

8 Man ville jämföra tiderna för två metoder (A och B) att producera en viss komponent i en bilmotor. Med metod A gjordes $n_A = 6$ mätningar och man erhöll medelvärdet $\bar{x}_A = 17.33$ och standardavvikelsen $s_A = 4.972$. Tidsenhet: minuter. Med metod B gjordes $n_B = 7$ mätningar och man erhöll medelvärdet $\bar{x}_B = 24.33$ och standardavvikelsen $s_B = 1.900$. Vi antar att man har verifierat (t.ex medelst kvantilplottar) att tiderna är approximativt normalfördelade. Det finns ingen anledning att tro att (de teoretiska) varianserna i de två stickproven är lika. Beräkna ett 99% konfidensintervall för differensen $\mu_B - \mu_A$ av väntevärdena för metoderna. Lyckades man (med konfidensen 99%) säkerställa att $\mu_A \neq \mu_B$? (3 p)

Lycka till med lösandet av uppgifterna!

Svar till Mat stat Z den 14/1-10

1. (a) 15 (b) 20
2. $3/35 \approx 0.086$
3. Poi(λ), $\hat{\lambda} = 1.4$, $\hat{p} = 0.753$, jfr $4/5 = 0.8$
4. 0.97
5. Notera först att överlevnadsfunktionen är $R(x) = e^{-Z(x)}$, där $Z(x) = \int_0^x z(y)\,dy = (c/1.1)x^{1.1}$. Det följer att tätheten är

$$f(x) = -\frac{d}{dx}\,R(x) = cx^{0.1}e^{-(c/1.1)x^{1.1}}$$

Således gäller att trolighetsfunktionen är

$$L(c) = \prod_i f(x_i) = c^n\,(x_1 \cdots x_n)^{0.1}\,e^{-(c/1.1)\sum_i x_i^{1.1}}$$

Efter logaritmering har vi

$$\mathcal{L}(c) = n\log c - (c/1.1)\sum_i x_i^{1.1} + \text{något som ej beror av } c$$

Vi får formeln för ML-skattningen genom att lösa ut c ur ekvationen

$$\frac{d}{dc}\,\mathcal{L}(c) = 0 \quad \Leftrightarrow \quad \frac{n}{c} = \frac{\sum_i x_i^{1.1}}{1.1}$$

Alltså,

$$\hat{c} = \frac{1.1n}{\sum_i x_i^{1.1}} = \frac{1.1 \cdot 5}{1.25 \cdot 10^8} = 44 \cdot 10^{-9}$$

6. (a) Nej (b) $H_0 : \mu = 100$ (eller $H_0 : \mu \le 100$) mot $H_1 : \mu > 100$. Teststatistikan, som är $t = \frac{\bar{x}-100}{s/\sqrt{n}} = 2.166$, har P-värdet 0.024 ($t_{0.025} = 2.145 < t < t_{0.01} = 2.624$). Vi kan alltså förkasta på 5%-nivån, men ej på 1%-nivån.
7. $p_X = 0.526 \pm 0.025 \in (0.501,\, 0.551)$
8. $\mu_B - \mu_A = 7.00 \pm 7.98 \in (-0.98,\, 14.98)$; Nej, man lyckades inte säkerställa att $\mu_A \neq \mu_B$ med konfidensen 99%.