



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I 5B1508 MATEMATISK STATISTIK FÖR S
TISDAGEN DEN 20 DECEMBER 2005 KL 08.00–13.00.

Examinator: Gunnar Englund, tel. 790 7416.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik. Räknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng.

Den som fått 18 eller 19 poäng på tentamen har möjlighet till komplettering. Kompletteringen skall göras inom tre veckor efter det att resultatet av denna tentamen publicerats. Den som är aktuell för komplettering skall till examinator anmäla önskan att få en sådan inom två veckor från publicering av tentamensresultatet.

Den som fått godkänt på lappskrivning nummer 1 från den 16 november 2005, får uppgift 1 a) tillgodoräknad. Den som fått godkänt på lappskrivning nummer 2 från den 30 november 2005, får uppgift 2 a) tillgodoräknad. Tillgodoräknade uppgifter skall inte lösas.

Resultatet anslås senast fredagen den 13 januari 2005 på Matematisk statistiks anslagstavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten.

Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivningstillfället.

Uppgift 1

a) För de oberoende händelserna A , B och C gäller att $P(A) = 0.3$, $P(B) = 0.4$ och $P(C) = 0.1$. Beräkna sannolikheten att åtminstone en av händelserna A , B och C inträffar. (6 p)

b) En hiss på KTH är märkt "Hiss för högst 400 kg eller högst 5 personer". En slumpmässigt vald manlig teknologs vikt betraktas som en $N(75, 10)$ -fördelad stokastisk variabel, och en slumpmässigt vald kvinnlig teknologs vikt som en $N(65, 5)$ -fördelad stokastisk variabel. Antag att 3 manliga och 2 kvinnliga teknologer tänker åka i hissen. Alla personernas vikter antas vara oberoende. Beräkna sannolikheten att teknologernas vikt överstiger 400 kg. (4 p)

Uppgift 2

En viss typ av mikrokretsar levereras i partier om 1000 enheter. Man använder följande provtagningschema: Ur varje parti väljs slumpmässigt utan återläggning ett urval om 80 enheter som kontrolleras. Om högst tre enheter i urvalet är defekta godkänns partiet – i annat fall underkänns det. Beräkna

a) sannolikheten att man underkänner ett parti där bråkdelen defekta enheter är 0.005. (6 p)

b) sannolikheten att man godkänner ett parti där bråkdelen defekta enheter är 0.15. (4 p)

Lämpliga och väl motiverade approximationer får göras i både a- och b-delen.

Uppgift 3

För prediktion av skjuvhållfasthet hos stålbalkar används i huvudsak en av två metoder, här kallade metod 1 och metod 2. I ett jämförande studium kontrollerade man kvoten mellan den predikterade belastningen mot den faktiska för fem olika balkar med de respektive metoderna. De erhållna kvoterna är som följer:

Stålbalk nr	1	2	3	4	5
Metod 1	1.19	1.15	1.32	1.34	1.20
Metod 2	1.06	0.99	1.06	1.06	1.07

Resultat k , $k = 1, \dots, n$, från metod 1 betraktas som ett utfall av $N(\mu_k + \Delta, \sigma_1)$ och resultat k från metod 2 betraktas som utfall av $N(\mu_k, \sigma_2)$. Alla stokastiska variabler i modellen antas vara oberoende.

a) Ställ upp ett konfidensintervall för den systematiska skillnaden Δ . Använd konfidensgrad 95%. (7 p)

b) Testa hypotesen $H_0 : \Delta = 0.25$ mot hypotesen $H_1 : \Delta \neq 0.25$ på nivå 5%. Det skall av ditt test klart framgå ifall H_0 kan förkastas eller ej. (3 p)

Uppgift 4

I en uppsats i *Concrete Research*, Vol. 41, 1989, presenteras data för sambandet mellan tryckhållfasthet x och permeabilitet y för olika blandningar av betong.

En sammanfattning av dessa data, som omfattar $n = 14$ par (x_i, y_i) av observationer, är

$$\sum_{i=1}^{14} y_i = 572 \qquad \sum_{i=1}^{14} y_i^2 = 23530$$

$$\sum_{i=1}^{14} x_i = 43 \qquad \sum_{i=1}^{14} x_i^2 = 157.42$$

$$\sum_{i=1}^{14} x_i y_i = 1697.80.$$

Vi antar att $(x_1, y_1), \dots, (x_{14}, y_{14})$ är relaterade till varandra enligt en linjär regressionsmodell dvs. som utfall av $(x_1, Y_1), \dots, (x_{14}, Y_{14})$, där varje Y_i är normalfördelad $N(\alpha + \beta x_i, \sigma)$

a) Bestäm skattningarna α_{obs}^* , β_{obs}^* för koefficienterna i regressionslinjen. (3 p)

b) Använd den skattade regressionslinjen för att skatta det förväntade värdet på permeabilitet när tryckhållfastheten x är 4.3. (2 p)

d) Bestäm konfidensintervall med konfidensgrad 95% för α respektive β . (5 p)

Uppgift 5

För en bitström över en brusig kanal kan bitar (1:or och 0:or) överföras felaktigt. Antag att en bit (1:a eller 0:a) överförs felaktig med sannolikhet p oberoende av andra bitar. Över kanal A är felsannolikheten för en bit 0.005 och över kanal B är felsannolikheten 0.015.

När man mottog en bitström om $n = 1000$ bitar såg man att $x = 10$ bitar var felaktigt överförda.

a) Skatta med Maximum Likelihood-metoden vilken kanal (A eller B) det var som överförde bitströmen. (5 p)

b) Man vet att 75% av alla bitströmar går över kanal A, resten över kanal B. Med

$$A = \{\text{Kanal A används}\} \quad \text{och} \quad B = \{\text{Kanal B används}\} = A^*$$

är $P(A) = 0.75$ och $P(B) = 0.25$.

Beräkna $P(A|10 \text{ fel av } 1000 \text{ bitar})$ och $P(B|10 \text{ fel av } 1000 \text{ bitar})$ och skatta använd kanal med hjälp av dessa betingade sannolikheter. (5 p)

Ledning: Skattningen blir en så kallad *Bayes*-skattning.



Avd. Matematisk statistik

KTH Matematik

FÖRSLAG TILL LÖSNINGAR

TENTAMEN I 5B1508 MATEMATISK STATISTIK FÖR S 2005-12-20

Uppgift 1

a) Vi skall beräkna

$$\begin{aligned} P(A \cup B \cup C) &= 1 - P(A^* \cap B^* \cap C^*) = (\text{oberoende}) = 1 - P(A^*)P(B^*)P(C^*) = \\ &= 1 - (1 - 0.3)(1 - 0.4)(1 - 0.1) = \underline{0.6220}. \end{aligned}$$

b) Låt X_i = 'vikten av manlig teknolog nr i ', $i = 1, 2, 3$ och Y_i = 'vikten av kvinnlig teknolog nr i ', $i = 1, 2$. Totala vikten är $S = X_1 + X_2 + X_3 + Y_1 + Y_2$ som är normalfördelade eftersom den är summan av oberoende normalfördelade stokastiska variabler. Vi får $E(S) = E(X_1) + E(X_2) + E(X_3) + E(Y_1) + E(Y_2) = 3 \cdot 75 + 2 \cdot 65 = 355$ och $V(S) = (\text{oberoendet}) = V(X_1) + V(X_2) + V(X_3) + V(Y_1) + V(Y_2) = 3 \cdot 10^2 + 2 \cdot 5^2 = 350$, dvs S är $N(355, \sqrt{350})$. Vi söker

$$P(S > 400) = P\left(\frac{S - 355}{\sqrt{350}} > \frac{400 - 355}{\sqrt{350}}\right) = 1 - \Phi\left(\frac{45}{\sqrt{350}}\right) \approx 1 - \Phi(2.41) \approx \underline{0.0080}.$$

Uppgift 2

X = antalet defekta i urvalet.

a) X är Hyp(1000, 80, 0.005)-fördelad. Då $n/N = 80/1000 = 0.08 \leq 0.1$ så är X approximativt Bin(80, 0.005)-fördelad.

Då $p = 0.005 \leq 0.1$ så är X approximativt Po($80 \cdot 0.005$) = Po(0.4)-fördelad. Således fås $P(X > 3) = 1 - P(X \leq 3) = (\text{Poisson-tabellen}) \approx 1 - 0.99922 = \underline{0.00078}$.

b) X är Hyp(1000, 80, 0.15)-fördelad. Som ovan fås att X approximativt Bin(80, 0.15)-fördelad.

Då $np(1 - p) = 80 \cdot 0.15 \cdot 0.85 = 10.2 \geq 10$ så är X approximativt $N(80 \cdot 0.15, \sqrt{10.2}) = N(12, \sqrt{10.2})$ -fördelad. Således fås

$$\begin{aligned} P(X \leq 3) &= P\left(\frac{X - 12}{\sqrt{10.2}} \leq \frac{3 - 12}{\sqrt{10.2}}\right) = \Phi\left(-\frac{9}{\sqrt{10.2}}\right) \\ &= 1 - \Phi\left(\frac{9}{\sqrt{10.2}}\right) = 1 - \Phi(2.82) = 1 - 0.9976 = \underline{0.0024}. \end{aligned}$$

Uppgift 3

De parvisa skillnaderna

Stålbalk nr	1	2	3	4	5
y_i : (Metod 1 – Metod 2)	0.13	0.16	0.26	0.28	0.13

är utfall av $N(\Delta, \sigma)$ -fördelade stokastiska variabler, där Δ skattas med $\bar{y} = 0.192$ och σ med $s = 0.072595$. Då $\bar{y}$ är ett utfall av $\bar{Y}$, $\bar{Y}$ är $N(\Delta, \sigma/\sqrt{n})$, så är

$$\frac{\bar{Y} - \Delta}{S/\sqrt{n}} \quad t(n-1)\text{-fördelad.}$$

Ur $t(n-1) = t(4)$ -tabell fås $t_{0.025} = 2.78$ så ett 95% konfidensintervall för Δ blir

$$\Delta \in \bar{y} \pm t_{0.025} \frac{s}{\sqrt{n}} = 0.192 \pm 2.78 \cdot 0.0325 = \underline{0.192 \pm 0.090} = \underline{(0.10, 0.28)}.$$

b) Eftersom $\Delta = 0.25$ inte är ett orimligt värde på Δ enligt konfidensintervallet kan H_0 inte förkastas på nivå 5%.

Uppgift 4

a) Vi får $\bar{x} = 43/14$ samt $\bar{y} = 572/14$. Vidare är (med beteckningar från formelsamlingen)

$$\begin{aligned} S_{xx} &= \sum_{i=1}^{14} x_i^2 - 14(\bar{x})^2 = 157.42 - 14 \cdot (43/14)^2 = 25.3486 \\ S_{xy} &= \sum_{i=1}^{14} x_i y_i - 14\bar{x}\bar{y} = 1697.80 - 14 \cdot (43/14) \cdot (572/14) = -59.0571 \\ S_{yy} &= \sum_{i=1}^{14} y_i^2 - 14(\bar{y})^2 = 23530 - 14 \cdot (572/14)^2 = 159.7143 \end{aligned}$$

Detta ger

$$\beta_{\text{obs}}^* = S_{xy}/S_{xx} = -2.3298, \quad \alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x} = 572/14 - (-2.3298) \cdot (43/14) = 48.0130.$$

Skattad linje är alltså $y = 48.0130 - 2.3298x$.

b) Vi skattar med $\alpha_{\text{obs}}^* + \beta_{\text{obs}}^* 4.3 = 48.0130 - 2.3298 \cdot 4.3 = 37.9949$.

c) Vi skattar σ med s där

$$(14-2)s^2 = S_{yy} - (\beta_{\text{obs}}^*)^2 S_{xx} = 22.1229$$

som ger $s = \sqrt{22.1229/12} = 1.3578$. Vi får konfidensintervallen

$$\begin{aligned} I_\alpha &= \alpha_{\text{obs}}^* \pm t_{0.025}(12) s \sqrt{\frac{1}{14} + \frac{(\bar{x})^2}{S_{xx}}} = 48.0130 \pm 2.18 \cdot 1.3578 \sqrt{\frac{1}{14} + \frac{(43/14)^2}{25.3486}} \\ &= \underline{48.0130 \pm 1.9714}, \\ I_\beta &= \beta_{\text{obs}}^* \pm t_{0.025}(12) \frac{s}{\sqrt{S_{xx}}} = \underline{-2.3298 \pm 0.5879}. \end{aligned}$$

Uppgift 5

Antalet felaktigt överförda bitar, $x = 10$, är ett utfall av en binomialfördelad stokastisk variabel X , X är $\text{Bin}(n, p) = \text{Bin}(1000, p)$.

Likelihoodfunktionen blir

$$\begin{aligned} L(p) &= P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \\ &= \begin{cases} \binom{1000}{10} 0.005^{10} (1-0.005)^{990} = 0.018 & \text{om } p = 0.005 \\ \binom{1000}{10} 0.015^{10} (1-0.015)^{990} = 0.048 & \text{om } p = 0.015 \end{cases} \end{aligned}$$

Då $L(0.015) > L(0.005)$ är ML-skattningen att kanal B användes. (Man kan använda Poissonapproximation och tabell 5 för att bestämma sannolikheterna. Då får man att $L(0.015) = 0.049$ och $L(0.005) = 0.018$.)

b) Med Bayes formel fås

$$\begin{aligned} P(A|X = 10) &= \frac{P(A \cap \{X = 10\})}{P(X = 10)} = \frac{P(X = 10|A)P(A)}{P(X = 10|A)P(A) + P(X = 10|B)P(B)} \\ &= \frac{L(0.005)P(A)}{L(0.005)P(A) + L(0.015)P(B)} = 0.528 \end{aligned}$$

och $P(B|X = 10) = 1 - P(A|X = 10) = 0.472$. Här är $P(A|X = 10) > P(B|X = 10)$ så skattningen blir att kanal A användes.