TENTAMEN I 5B1507 MATEMATISK STATISTIK FÖR T och M TORSDAGEN DEN 12 JANUARI 2006 KL 08.00–13.00.

Examinator: Jan Enger, tel. 790 7134.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik. Formelblad statistisk kvalitetsstyrning. 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 24 poäng.

Den som fått 22 eller 23 poäng på tentamen har möjlighet till komplettering. Kompletteringen skall göras inom två 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 en vecka från publicering av tentamensresultatet.

Den som fått godkänt på lappskrivning nummer 1 från den 23 september 2005, får uppgift 1 a) tillgodoräknad. Den som fått godkänt på lappskrivning nummer 2 från den 5 oktober 2005, får uppgift 4 a) tillgodoräknad. Tillgodoräknade uppgifter skall inte lösas.

Resultatet anslås senast torsdagen den 2 februari 2006 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) En diskret stokastisk variabel X kan anta värdena 1, 2, 3, 4 och 5 och har sannolikhetsfunktion

k	1	2	3	4	5
$p_X(k)$	0.1	0.2	0.2	0.3	?

Beräkna $p_X(5)$, väntevärdet $E(2X-3)$ och standardavvikelsen $D(2X-3)$. (5 p)

b) I en process blir enheter felaktiga oberoende av varandra med sannolikhet p. Om två enheter av i följd tillverkade fem blir felaktiga, justeras processen.

Vad är sannolikheten att processen justeras för första gången då den sjunde enheten tillverkats. Bestäm även det numeriska värdet på denna sannolikhet om $p = 0.1$. (5 p)

Uppgift 2

a) Låt X vara slumpmässigt vald i intervallet $(0\,,3)$ och antag att Y är, oberoende av X, slumpmässigt valt i intervallet $(0\,,4)$, dvs att $X \in U(0,3)$ och $Y \in U(0,4)$.
Beräkna $P(X \geq Y)$. (5 p)

b) Antag att $X_1, X_2, \ldots, X_{25}$ alla är $U(0,3)$ och $Y_1, Y_2, \ldots, Y_{25}$ alla är $U(0,4)$. Alla variabler antas oberoende av varandra.
Beräkna $P(\sum_{i=1}^{25} X_i \geq \sum_{i=1}^{25} Y_i)$. Välmotiverade approximationer är tillåtna. (5 p)

Uppgift 3

I en produktionsprocess varierar diametern av tillverkade kulor som $N(25.000, 0.020)$ mm. En kula vars diameter ligger utanför intervallet (24.930, 25.070) kasseras.

a) Vad är sannolikheten att en tillverkad kula kasseras. (4 p)

b) Av 10000 tillverkade kulor, vad är sannolikheten att högst 5 stycken kasseras. Välmotiverade approximationer är tillåtna. (6 p)

Uppgift 4

a) Ett voltmeter har ett systematiskt fel 0.5 mV och ett slumpmässigt fel som är $N(0, 0.03)$ mV, varför en mätning av en spänning på μ mV ger ett mätvärde som är $N(\mu + 0.5, 0.03)$. Vid 6 mätningar av en okänd spänning erhölls resultaten (i mV)

10.53, 10.55, 10.55, 10.56, 10.57 och 10.57

Ge ett 95 % konfidensintervall för den okända spänningen, samt testa på 5 % signifikansnivå hypotesen att den okända spänningen är 10.00. (5 p)

b) Med ett annat mätinstrument mättes samma spänning och man erhöll resultaten

9.99, 10.04, 10.04 och 10.05.

Oberoende, normalfördelade observationer med samma standardavvikelsen som för instrumentet i a).
Testa med, hjälp av de två stickproven, på lämpligt sätt hypotesen att det andra instrumentet saknar systematiskt fel (dvs att det är 0). Använd 5 % signifikansnivå. (5 p)

Uppgift 5

a) Dämpningen av laserstyrkan för ett laserinstrument som används för avståndsmätningar uppmättes för avstånden 600, 800 och 1000 meter med följande resultat (där också medelvärden för dämpningarna vid de olika avstånden angivits).

Avstånd (x)	Observationer (y)					medelvärde
600	10.1	12.1	10.7	11.5		11.1
800	15.3	16.7	15.5	17.1	16.9	16.3
1000	20.5	22.6	21.3	22.0		21.6

Hjälpsummor: $\sum_i (y_i - \bar{y})^2 = 228.0877, \quad \bar{y} = 16.3308$

a) Anpassa en linjär regressionsmodell $Y = \alpha + \beta x + \varepsilon$, $\varepsilon \in N(0, \sigma)$, och skatta parametrarna α, β och σ^2 i denna. (5 p)

b) Ge ett 95 % konfidensintervall för förväntad dämpning vid avståndet 900 meter. (3 p)

c) Ge ett 95 % konfidensintervall för regressionslinjens lutning. (2 p)

Uppgift 6

För att undersöka en bilmodells bromseffekt mättes bromssträckan efter inbromsning vid hastigheterna 20, 40, 60, 80 och 100 km/h. Resultat:

hastighet v (20-tal km/h)	1	2	3	4	5
bromssträcka x (meter)	12	25	52	78	119

a) Rent teoretiskt skulle bromssträckan vara en kvadratisk funktion av hastigheten plus en normalfördelad slumpmässig avvikelse, och därför antas bromssträckan vid inbromsning vid hastigheten v vara $\mathrm{N}(\gamma v^2, \sigma^2)$ där γ är okänd. Skatta γ med minsta kvadrat-metoden. (4 p)

b) Undersök om MK-skattningen i a) är väntevärdesriktig. (3 p)

c) Beräkna variansen för MK-skattningen i a) (uttryckt i σ). (3 p) figure=./figurer/kth$_matematik_cmyk.eps, height = 30mm$

FÖRSLAG TILL LÖSNINGAR
TENTAMEN I 5B1507 MATEMATISK STATISTIK FÖR T och M 2006–01–12

Uppgift 1

a) Eftersom $\sum_{k=1}^{5} p_X(k) = 1$ fås $p_X(5) = 0.2$. Härav fås att $E(X) = 1 \cdot 0.1 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.3 + 5 \cdot 0.2 = 3.3$, $E(X^2) = 1 \cdot 0.1 + 4 \cdot 0.2 + 9 \cdot 0.2 + 16 \cdot 0.3 + 25 \cdot 0.2 = 12.5$ och därav $V(X) = E(X^2) - (E(X))^2 = 1.61$. Det ger oss $E(2X - 3) = 2E(X) - 3 = 3.6$, $V(2X - 3) = 4V(X) = 6.44$ och därmed $D(2X - 3) = \sqrt{6.44} = 2.54$.

b) Att processen skall justeras för första gången då den sjunde enheten tillverkats innebär att ett av två följande fall inträffat.

- Den första, sjätte och sjunde enheten är felaktiga övriga korrekta.
- Enhet ett och två korrekta, exakt en av enheterna tre till sex felaktig och den sjunde felaktig.

Sannolikheten för det första fallet är $p^3 \cdot (1 - p)^4$.
Antal enheter av enheterna tre till sex som är felaktiga är $\text{Bin}(4, p)$ och därför är sannolikheten att exakt en av dessa felaktiga lika med $\binom{4}{1}p(1 - p)^3 = 4p(1 - p)^3$. Sannolikheten för fall 2 är därför $(1-p)^2 \cdot 4 \cdot p(1-p)^3 \cdot p = 4p^2(1-p)^5$. Summan av sannolikheterna för dessa två fall är den sökta sannolikheten, $p^2(1 - p)^4(4 - 3p)$. För $p = 0.1$ fås värdet 0.0243.

Uppgift 2

a) Vi har att

$$P(X \geq Y) = \iint_{x\geq y} f_{X,Y}(x, y)dxdy = \iint_{x\geq y} f_X(x)f_Y(y)dxdy = \frac{1}{12} \iint\limits_{\substack{x\geq y\\0\leq x\leq 3\\0\leq y\leq 4}} dxdy$$

Integralens värde är arean av det skuggade området i figuren som är 4.5. Den sökta sannolikheten är alltså $P(X \geq Y) = \frac{4.5}{12} = \frac{3}{8}$.

figure=./figurer/0601-t.eps,height=40mm

b) Av centrala gränsvärdessatsen följer att $\sum_{i=1}^{25} X_i$ är approximativt $N(25{\cdot}1.5, \sqrt{25 \cdot 3^2/12}) = N(37.5, \sqrt{225/12})$ eftersom X_i:na har väntevärde 1.5 och varians $3^2/12$. På samma sätt är $\sum_{i=1}^{25} Y_i$ approximativt $N(50, \sqrt{400/12})$. Vi får därför att $\sum_{i=1}^{25} Y_i - \sum_{i=1}^{25} X_i$ är approximativt $N(50 - 37.5, \sqrt{225/12 + 400/12}) = N(12.5, 25/\sqrt{12})$. Det ger

$$P(\sum_{i=1}^{25} X_i \geq \sum_{i=1}^{25} Y_i) = P(\sum_{i=1}^{25} Y_i - \sum_{i=1}^{25} X_i \leq 0) \approx \Phi(\frac{0 - 12.5}{25/\sqrt{12}}) = \Phi(-\sqrt{3}) = 1 - \Phi(\sqrt{3}) \approx 0.042.$$

Uppgift 3

a) Låt X vara en kulas diameter. Vi får att $P(24.93 \leq X \leq 25.07) = \Phi(\frac{25.07-25.00}{0.02}) - \Phi(\frac{24.93-25.00}{0.02}) = 2\Phi(\frac{0.07}{0.02}) - 1 = 0.99954$. Sökt sannolikhet: $1 - 0.99954 = 0.00046$.

b) Låt Y vara antal kasserade kulor. Då är $Y \in$Bin$(10000, 0.00046)$ som är approximativt Po(4.6) eftersom $p = 0.00046 < 0.10$. Tabell ger nu att $P(Y \leq 5) = 0.686$

Uppgift 4

a) Ett 95 % konfidensintervall för väntevärdet μ+0.5 ges av $\bar{x}\pm\lambda_{0.025}\frac{\sigma}{\sqrt{n}} = 10.555\pm 1.96\frac{0.03}{\sqrt{6}} = 10.555\pm 0.024$ (se FS 11.1a)+12.1). Ett 95 % konfidensintervall för μ blir därför 10.055 ± 0.024. Värdet 10.00 ingår inte i intervallet och hypotesen $\mu = 10.00$ förkastas därför.

b) Observationerna är $N(\mu + \Delta, 0.03)$ där Δ är det systematiska felet. Skillnaden mellan de förväntade värdena i b) och a) är således $\mu + \Delta - (\mu + 0.5) = \Delta - 0.5$. Ett 95 % konfidensintervall för denna väntevärdesskillnad ges av $\bar{y} - \bar{x} \pm 1.96 \cdot 0.03 \cdot \sqrt{1/6 + 1/4} = 10.03 - 10.555 \pm 1.96 \cdot 0.03\sqrt{5/12} = -0.525 \pm 0.038$ (se FS 11.2a)+12.1). Ett 95 % konfidensintervall för Δ ges därför av -0.025 ± 0.038. Värdet 0 ingår i intervallet och hypotesen att systematiskt fel saknas kan inte förkastas på 5 % signifikansnivå.

Uppgift 5

a) Vi har $\bar{x} = \frac{4\cdot 600+5\cdot 800+4\cdot 1000}{13} = 800$ och
$\sum_i (x_i - \bar{x})^2 = 4 \cdot (600 - 800)^2 + 5 \cdot (800 - 800)^2 + 4 \cdot (1000 - 800)^2 = 320000$. Vidare gäller $\sum_i (x_i - \bar{x})y_i = -200 \cdot (10.1 + 12.1 + 10.7 + 11.5) + 0 + 200 \cdot (20.5 + 22.6 + 21.3 + 22.0) = 8400$ och därför får vi enligt FS 13.1 och 13.3

$$\beta^* = \frac{8400}{320000} = 0.02625, \quad \alpha^* = \bar{y} - \beta^*\bar{x} = -4.669 \text{ och } s^2 = \frac{228.0877 - 8400^2/320000}{13 - 2} = 0.690.$$

b) Enligt FS 13.2 ges konfidensintervallet av $\alpha^* + \beta^* \cdot 900 \pm t_{0.025}(13 - 2)s\sqrt{\frac{1}{13} + \frac{(900-800)^2}{320000}} = 18.96 \pm 0.60$ eftersom $t_{0.025}(11) = 2.20$.

c) Ett konfidensintervall för β ges av $\beta^* \pm t_{0.025}(13 - 2)s/\sqrt{320000} = 0.0263 \pm 0.0032$.

Uppgift 6

a) Vi skall minimera $Q(\gamma) = (x_1 - \gamma)^2 + (x_2 - 2^2\gamma)^2 + (x_3 - 3^2\gamma)^2 + (x_4 - 4^2\gamma)^2 + (x_5 - 5^2\gamma)^2$.
Derivera; $Q' = -2(x_1 - \gamma + 4(x_2 - 4\gamma) + 9(x_3 - 9\gamma) + 16(x_4 - 16\gamma) + 25(x_5 - 25\gamma))$.
$Q' = 0$ ger $\gamma^*_{obs} = \frac{x_1+4x_2+9x_3+16x_4+25x_5}{979}$. Med värden insatta fås $\gamma^*_{obs} = 4.906$.

b) $E(\gamma^*) = E(\frac{X_1+4X_2+9X_3+16X_4+25X_5}{979}) = \frac{1}{979}(\gamma+4^2\gamma+9^2\gamma+16^2\gamma+25^2\gamma) = \gamma$. Väntevärdesriktig.

c) $V(\gamma^*) = V(\frac{X_1+4X_2+9X_3+16X_4+25X_5}{979}) = \frac{1}{979^2} \cdot (1 + 4^2 + 9^2 + 16^2 + 25^2) \cdot \sigma^2 = \sigma^2/979$.