

Korrekt och väl motiverad lösning på uppgifterna 1–5 ger 10 poäng vardera medan delfrågorna på uppgift 6 ger 4 poäng vardera. Totalt kan man få 70 poäng. Gränsen för godkänd är 35 poäng, dock finns det vissa minimikrav på uppgifterna 1–5 (18p) respektive uppgift 6 (7p).

Institutionens papper används både som kladdpapper och inskrivningspapper. Varje lösning skall börja överst på nytt papper. Rödpenna får ej användas.

Tillåtna hjälpmedel: Matematiska och statistiska tabeller som ej innehåller statistiska formler, formelsamling matematisk statistik för bio- och kemitekniker, samt miniräknare.

Resultatet förs in i LADOK senast måndag den 17 september.

Redovisa införda beteckningar; ange modeller, approximationer, hypoteser och slutsatser. Motivera alla antagande.

Skriv anonymkoden och identifierare (eller presonummer och namn) på omslaget och SAMTLIGA inlämnade papper.

1. För att uppskatta hur många valsedlar som behövs i en vallokal antas att varje väljare tar en valsedel från parti A med 75% sannolikhet (en del tar valsedlar från alla partier, en del från ett urval av partier och en del från endast det parti de tänker rösta på). Vallokalen har 1 200 röstberättigade som vi antar agerar oberoende av varandra.
 - (a) Om alla 1 200 väljare kommer, vad är då fördelningen (exakt) för antalet valsedlar från parti A som kommer att tas under valet. (3p)
 - (b) Ange och motivera en lämplig approximation av fördelningen i a). (3p)
 - (c) Hur många valsedlar från parti A behövs approximativt om det ska vara mindre än 1% chans att valsedlarna tar slut? (4p)
2. I Sifos senaste opinionsundersökning (13–16 augusti) frågades 3 486 personer om hur de tänker rösta i höstens riksdagsval.

Parti	Andel svarande (%)
Moderaterna	19.4
Liberalerna	5.1
Centerpartiet	9.5
Kristdemokraterna	4.2
Miljöpartiet	6.5
Socialdemokraterna	25.2
Vänsterpartiet	9.3
Sverigedemokraterna	18.7

- (a) Under antagande om oberoende och att de tillfrågade personerna är representativt utvalda, formulera en lämplig modell för antalet Kristdemokratiska väljare i undersökningen och ange ett approximativt 95%-igt konfidensintervall för andelen röster som partiet kan förväntas få i valet. (6p)
- (b) Varför är det viktigt att personerna är representativt utvalda? (2p)
- (c) Vad kan man säga om Kristdemokraternas chanser att komma in i riksdagen (d.v.s. få mer än 4% av rösterna)? (2p)

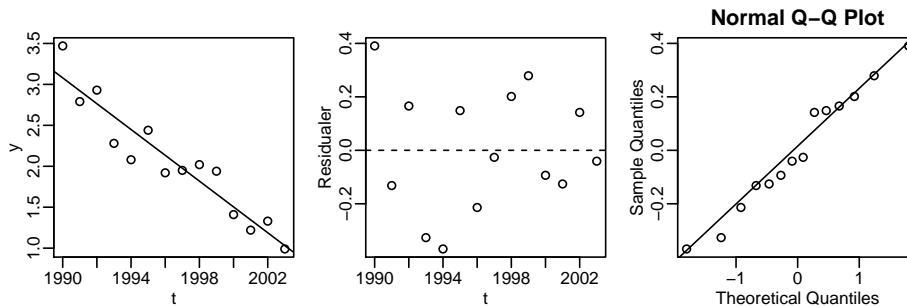
3. Ett flertal storstäder, t.ex. Paris, har experimenterat med restriktioner av (privat) biltrafik. En vanlig lösning är att endast bilar som har registreringsnummer med jämna eller udda slutsiffror för köra på jämna respektive udda dagar. För att undersöka effekterna av restriktionerna mätes dygnshalter av NO_2 ($\mu\text{g}/\text{m}^3$) för dagar utan (x_i) och med (y_i) restriktioner.

$$\begin{array}{lll} \bar{x} = 70.67 & s_x^2 = 11.32 & n_x = 100 \\ \bar{y} = 66.08 & s_y^2 = 11.52 & n_y = 90 \end{array}$$

Formulera en lämplig modell och avgör (på signifikansnivån 5%) om restriktionerna har sänkt halterna av NO_2 . Observationerna kan antas vara oberoende och normalfördelade med samma varians. (10p)

4. Bensen sprids i luften från bland annat bensindrivna fordon. Efter att ha infört katalytisk avgasrening och striktare regler för hur mycket bensen som får användas i bensinen har bensenhalten i luft, $\mu\text{g}/\text{m}^3$, i tätortsmiljö sjunkit sedan början av 1990-talet. Data är hämtade från Naturvårdsverket.

t_i , År	1990	1991	1992	1993	1994	1995	1996
y_i , Bensen $\mu\text{g}/\text{m}^3$	3.47	2.79	2.93	2.28	2.08	2.44	1.92
t_i , År	1997	1998	1999	2000	2001	2002	2003
y_i , Bensen $\mu\text{g}/\text{m}^3$	1.95	2.02	1.94	1.41	1.22	1.33	0.99



$$\begin{array}{lll} n = 14 & \bar{t} = 1996.5 & \bar{y} = 2.055 \\ S_{tt} = 227.5 & S_{ty} = -38.86 & S_{yy} = 6.3244 \end{array}$$

- Undersök om det är lämpligt att ansätta en linjär regressionsmodell $y_i = \alpha + \beta \cdot t_i + \varepsilon_i$, $\varepsilon_i \in N(0, \sigma)$, för hur bensenhalten varierar med tiden. (2p)
- Under förutsättning att en linjär regressionsmodell är lämplig, skatta samtliga parametrar i modellen. (2p)
- Avgör på lämpligt sätt om förändringen av bensenhalten med avseende på tiden är signifikant. (3p)
- Utifrån dessa data, vad är ett troligt värde på uppmätt bensenhalt år 2008? Svara med ett lämpligt intervall. (3p)

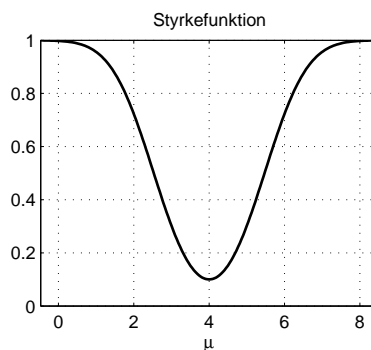
5. Kemisterna i ett mejeri övervägde att byta våglängd i sin spektrofotometriska metod för fosfatbestämning. Dessförinnan gjorde de ett 2^2 -försök med den gamla och den nya våglängden och prov med två olika fosfathalter. De mätte absorbansen för två replikat per försökspunkt, med följande resultat:

		Faktor B,	fosfathalt (μg)
		100	400
Faktor A	700	0.05, 0.03	0.16, 0.16
(våglängd)	820	0.07, 0.09	0.30, 0.34

Skatta huvudeffekterna och samspelseffekten och avgör vilka av dessa som är signifikanta. (Absorbanserna förutsätts vara normalfördelade.) (10p)

6. Teorifrågor. Ge koncisa svar på nedanstående frågor.

- (a) Antag att X och Y är oberoende stokastiska variabler med $X \in \text{Bin}(4, 0.1)$ och $Y \in \text{Bin}(5, 0.1)$. Bestäm sannolikheten $P(X + Y \leq 1)$ (4p)
- (b) Givet n stycken oberoende stokastiska variabler X_1, X_2, \dots, X_n med samma varians $\mathbf{V}(X_i) = \sigma^2$.
- Vad är $\mathbf{V}(n \cdot X_1)$ och $\mathbf{V}(\sum_{i=1}^n X_i)$? (2p)
 - Ge en (enkel) förklaring till varför $\mathbf{V}(n \cdot X_1) > \mathbf{V}(\sum_{i=1}^n X_i)$. (2p)
- (c) Vid radioaktivtsönderfall beskrivs överlevnadstiden för en enskild atom av en exponentialfördelning, $T \in \text{Exp}(\mu)$ medan halveringstiden är den tid då 50% av materialet har sönderfallit (d.v.s. den tid som 50% av atomerna överlever). Bestäm sambandet mellan materialets halveringstiden, $\tau_{1/2}$, och μ . (4p)
- (d) Beskriv sambandet mellan **ensidig** hypotesprövning och konfidensintervall för väntevärdet hos en normalfördelning. Kom ihåg att förklara båda fallen för ensidiga hypoteser ($<$ och $>$). (4p)
- (e) Figuren nedan visar styrkefunktionen för ett hypotestest.
- Vilka hypoteser har testats? (2p)
 - Vad är testets signifikansnivå? (2p)



LYCKA TILL!