
Lärare och Jour: David Bolin, telefon 772 53 75.

Tillåtna hjälpmedel: Formelsamling och valfri miniräknare med tömda minnen.

Korrekt, väl motiverad lösning ger poängen som är indikerad i parentes vid vardera uppgift. Totalt kan man få 40 poäng och betygsgränserna för betyg 3, 4 och 5 är 16, 24 och 32 poäng.

1. Vid en undersökning på en högstadieskola frågade man alla elever om de någon gång hade druckit alkohol. Resultatet från undersökningen kan ses i följande tabell:

	Årskurs 7	Årskurs 8	Årskurs 9
Andel av eleverna på skolan	0.40	0.25	0.35
Andel i årskursen som druckit alkohol	0.15	0.32	0.45

Den första raden visar alltså andelen av det totala antalet elever som går i varje årskurs och den andra raden visar för varje årskurs hur många som druckit alkohol.

- (a) Var är sannolikheten att en slumpvis vald elev på skolan går i årskurs nio och har druckit alkohol? (1p)
- (b) Vad är sannolikheten att en slumpvis vald elev på skolan har druckit alkohol? (2p)
- (c) Man finner en slumpvis vald elev på skolan som har druckit alkohol. Vad är sannolikheten att hen går i årskurs 9? (2p)
2. För att undersöka hur temperatur och luftfuktighet påverkar hållbarheten hos en livsmedelssprodukt utfördes ett 2^2 faktorförsök med två replikat. Följande tabell visar resultaten från försöket, där mätningarna visar antalet timmar det tog tills att produkten hade blivit dålig.

Temperatur (C)	Luftfuktighet (%)	Mätningar (timmar)
3	40	232, 254
20	40	135, 131
3	85	182, 185
20	85	89, 98

- (a) Skatta huvud- och samspelseffekterna. Vilken faktor verkar ha störst påverkan på hållbarheten? (2p)
- (b) Skatta variansen hos mätfelet. (1p)
- (c) Testa vilka av effekterna som är signifikanta på nivå 5%. (2p)
3. Mängden Ibuprofen (mg) i värktabletter av ett visst märke antas vara $N(205, 10^2)$ -fördelad.
- (a) Vad är sannolikheten att mängden Ibuprofen i en värktablett är mindre än 200mg? (1p)
- (b) Tabletterna säljs i förpackningar med tio tabletter och tillverkaren skriver att den totala mängden Ibuprofen i förpackningen ska vara åtminstone 2g. Vad är sannolikheten för att detta stämmer? (2p)
- (c) Tillverkaren kan spara pengar under tillverkningen genom att tillåta en högre varians hos mängden Ibuprofen i tabletterna. Anta att mängden är $N(205, \sigma^2)$ -fördelad. Hur stor kan σ^2 väljas så att sannolikheten för att det är minst 2g Ibuprofen i en förpackning är åtminstone 0.9? (2p)

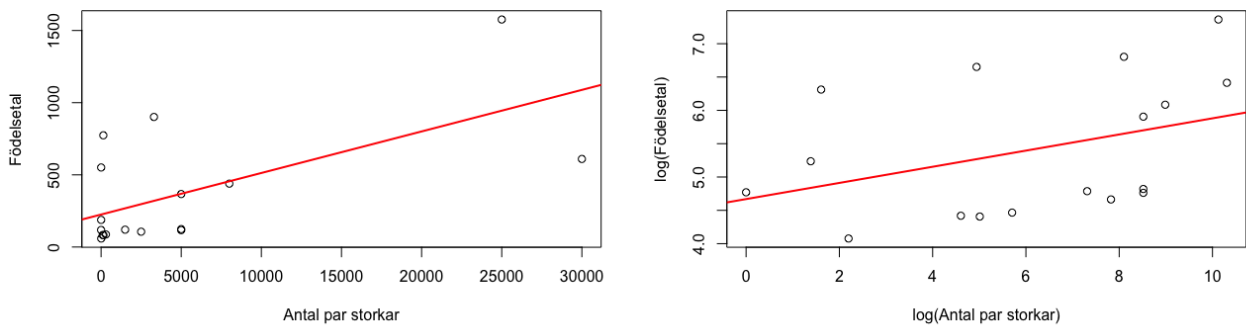


Figure 1: Till vänster visas data för födelsetal (enhet 1000 barn per år) och antal par storkar i 17 europeiska länder samt skattad regressionslinje. Till höger visas samma data i logaritmerad skala.

4. Ett uttryck som de flesta känner till är att storken kommer med barnen. Inte helt oväntat finns det vetenskapliga studier som undersöker detta påstående. I en sådan (Matthews, 2000) finns data för 17 länder i Europa, som visas till vänster i Figur 1. För att få uttryck som är enklare att hantera använder vi logaritmerade värden (naturliga logaritmen) för de två variablerna, och dessa visas till höger i figuren. Baserat på den logaritmerade datan kan vi beräkna $S_{xx} = 167.9901$, $S_{yy} = 16.6234$, $S_{xy} = 20.3658$, $\bar{x} = 6.0981$, $\bar{y} = 5.4084$.
- Låt X beteckna logaritmen av antalet par storkar och låt Y beteckna logaritmen av landets födelsetal. Skatta parametrarna i den linjära regressionsmodellen $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ där $\varepsilon_i \sim N(0, \sigma^2)$. (2p)
 - 2014 fanns 44 par storkar i Sverige. Gör ett prediktionsintervall för logaritmen av Sveriges födelsetal baserat på detta. Det föddes 113600 barn i Sverige 2014, ligger det observerade värdet i prediktionsintervallet? (1p)
 - Testa på nivå 5% om β_1 är signifikant. Vad är din slutsats? Om β_1 är signifikant, betyder detta att antalet storkar påverkar födelsetalet? Motivera svaret. (2p)
5. En tillverkare av vattenkokare skriver att uppkokningstiden för en liter vatten är tre minuter och 20 sekunder. För att testa om detta verkar stämma gjordes ett test då man kokade upp vatten från rumstemperatur tio gånger i rad och mätte tiden för varje försök. Resultatet ges i följande tabell.

Försök nummer	1	2	3	4	5	6	7	8	9	10
Tid (sekunder)	201	219	199	196	199	201	201	209	205	213

- Använd ett teckentest för att testa om mediankoktiden är signifikant större än 200 sekunder på nivå 5%. (3p)
- Antag att koktiderna är normalfördelade och utför ett t-test för att testa om medelkoktiden är signifikant större än 200 sekunder på nivå 5%. (3p)
- Vad är för och nackdelarna med de två testerna i (a) och (b)? (1p)
- Ett annat dyrare märke av vattenkokare hävdar att deras modell har den bästa koktiden på marknaden. För att testa detta utfördes samma experiment som ovan med deras modell, med följande resultat.

Försök nummer	1	2	3	4	5	6	7	8	9	10
Tid (sekunder)	191	208	185	209	198	202	180	187	212	220

Antag att mätningarna för båda modellerna är normalfördelade med samma varians. Testa om den dyrare modellen har en signifikant lägre medelkottid på nivå 5%. (3p)

6. Paretofördelningen är en vanlig kontinuerlig fördelning som bland annat har använts för att modellera hur den totala förmögenheten i en befolkning är allokerad mellan individerna. Låt (X_1, X_2) vara en tvådimensionell slumpvariabel med följande täthetsfunktion,

$$f(x_1, x_2) = \begin{cases} a(a+1)(x_1 + x_2 - 1)^{-(a+2)} & \text{om } x_1 > 1 \text{ och } x_2 > 1 \\ 0 & \text{för övrigt,} \end{cases}$$

där $a > 0$ är en parameter. Variabeln (X_1, X_2) sägs då ha en tvådimensionell Paretofördelning.

- (a) Härled marginalfördelningen för X_1 . (3p)
- (b) Beräkna $P(X_1 > 2)$. (2p)
- (c) För denna fördelning är väntevärdet endast definierat om $a > 1$. Om vi antar att $a > 1$, vad är då $E(X_1)$? (1p)
- (d) Vad är den betingade fördelningen för X_2 givet $X_1 = x_1$? (1p)
- (e) Antag att vi har ett stickprov med observationer av den första variabeln $X_1, x_{1,1}, \dots, x_{1,n}$. Beräkna maximum likelihood-skattaren av parametern a baserat på detta stickprov. (3p)

Lycka till!