Matematiska vetenskaper	**Tentamen: 2016-12-20 kl** $14^{00} - 18^{00}$
Chalmers tekniska högskola	**Matematisk statistik för K, TMA074**

Lärare: David Bolin, telefon 772 53 75. **Jour**: Sandra Eriksson Barman, telefon 772 53 77.

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. (a) Beräkna $\mathsf{P}(X > Y)$ om $X \sim \mathsf{N}(3, 2^2)$ och $Y \sim \mathsf{N}(2, 3^2)$. (2p)

   (b) Antag att X är en diskret slumpvariabel med fördelningsfunktion given av följande tabell

k	1	2	3	4	5
F(k)	0.3	0.4	0.7	0.8	1

   beräkna sannolikhetsfunktion, väntevärde och varians för X. (3p)

2. Man vill undersöka effekten av en viss medicin och mäter blodtrycket på sju patienter före och efter de tar medicinen. Antag att de två uppmätta värdena är normalfördelade enligt $N(\mu_i, \sigma^2)$ respektive $N(\mu_i + \Delta, \sigma^2)$. Vid mätningarna erhölls följande resultat av övertrycket:

Observation i	1	2	3	4	5	6	7
Före	101	110	99	130	170	125	144
Efter	104	100	90	133	150	120	130

   (a) Skatta Δ i modellen. (2p)

   (b) Beräkna ett tvåsidigt 95% konfidensintervall för Δ. (1p)

   (c) Testa $H_0 : \Delta = 0$ mot $H_1 : \Delta < 0$ på nivå $\alpha = 0.05$. (2p)

3. Vi har tio oberoende observationer $k_1, \ldots, k_{10}$ från en negativ binomialfördelning $\mathrm{NB}(r, p)$ med parameter $r = 3$.

   (a) Beräkna maximum likelihood-skattaren av p baserat på den observerade datan. (4p)

   (b) Binomialfördelningen $\mathrm{Bin}(n, p)$ används för att beskriva fördelningen för antalet lyckade försök bland n oberoende Bernoulliförsök som har sannolikhet p att lyckas. Vad är motsvarande tolkning för den negativa binomialfördelningen? (1p)

4. En hemsideutvecklare vill undersöka om färgen på en sida påverkar tiden som besökarna spenderar på sidan. Under fyra dagar testar hon fyra olika färger och mäter varje dag antalet besökare samt medeltiden bland besökarna.

Färg	Antal besökare	Medeltid
Vit	55	33s
Gul	47	27s
Grön	33	45s
Blå	68	40s

   Utvecklaren vill nu använda variansanalys för att undersöka om färgen påverkar längden.

   (a) Skriv ner modellen som antas för datan vid variansanalys. (1p)

   (b) Beräkna variansanalystabellen för försöket. Den totala kvadratsumman för observationerna ges av $SS_{tot} = \sum_{ij}(y_{ij} - \bar{y})^2 = 18000$. (3p)

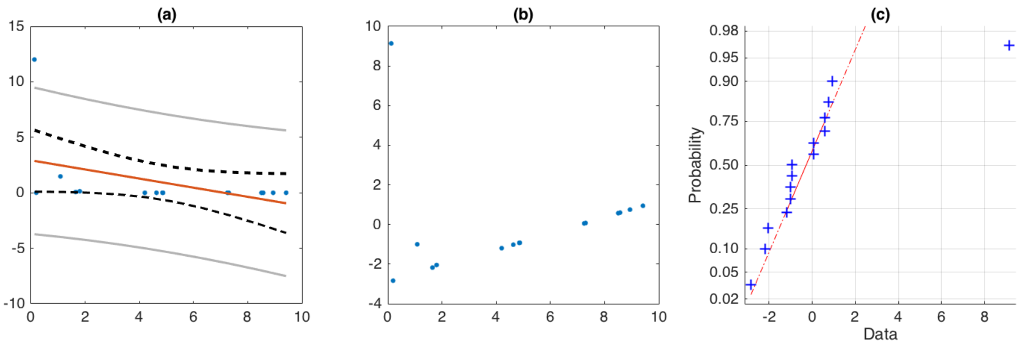

Figure 1: Resultat för Modell 1 i Uppgift 5. I (a) visas datan, skattad regressionslinje, konfidensintervall för linjens läge och prediktionsintervall. I (b) visas en plot av residualerna och i (c) visas en normalfördelningsplot av residualerna.

(c) Utför ett F-test för att undersöka om färgen påverkar längden. (2p)

5. Man vill modellera hur mottagen effekt hos en laptop påverkas av avståndet till den trådlösa routern i en lägenhet och gör därför 15 mätningar av mottagen effekt vid olika avstånd. För att beskriva sambandet testades två olika regressionsmodeller:

$$\begin{aligned} &\text{Modell 1:} & y_i &= \beta_0 + \beta_1 d_i + \varepsilon_i \\ &\text{Modell 2:} & \log(y_i) &= \beta_0 + \beta_1 d_i + \varepsilon_i \end{aligned}$$

där y_i är mottagen effekt, d_i är avstånd och $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$. Medelavståndet vid mätningarna var $\bar{d} = 4.89$ och spridningen i avstånden var $S_{xx} = 154.66$. Vid anpassning av de två modellerna erhölls följande resultat:

	Modell 1		Modell 2	
Parameter	skattning	95% konfidensintervall	skattning	95% konfidensintervall
β_0	2.92	(-0.01,5.85)	0.71	(-1.95,3.39)
β_1	-0.41	(-0.91,0.09)	-2.35	(-2.81,-1.90)
σ	2.77		2.53	

Förklaringsgraden för Modell 1 är $R^2 = 0.1949$ medan den är $R^2 = 0.90$ för Modell 2.

(a) Baserat på resultaten för Modell 1, kan man hävda att avstånd har en signifikant påverkan på mottagen effekt? Motivera svaret. (2p)

(b) Definiera förklaringsgraden R^2 för en linjär regressionsmodell. (2p)

(c) För att undersöka vilken av de två modellerna som passar bäst utförs en residualanalys, resultaten för modellerna kan ses i Figur 1 respektive Figur 2. Jämför resultaten och motivera vilken av de två modellerna som är bäst. (2p)

(d) Beräkna den förväntade logaritmerade effekten vid avstånd $d = 7$ baserat på Modell 2. Beräkna också ett 95% konfidensintervall för storheten. (2p)

(e) Svara på om följande påståenden är sanna eller falska, du behöver inte motivera dina svar. Korrekt svar ger 0.5 poäng, inget svar ger 0 poäng och felaktigt svar ger -0.5 poäng. Totalt kan du lägst få 0 poäng på deluppgiften. (2p)

i. De sträckade linjerna i Figur 1 anger konfidensintervallet för linjens läge.

ii. Konfidensintervallet för linjens läge skulle bli bredare om vi ökar konfidensgraden från 95% till 99%.

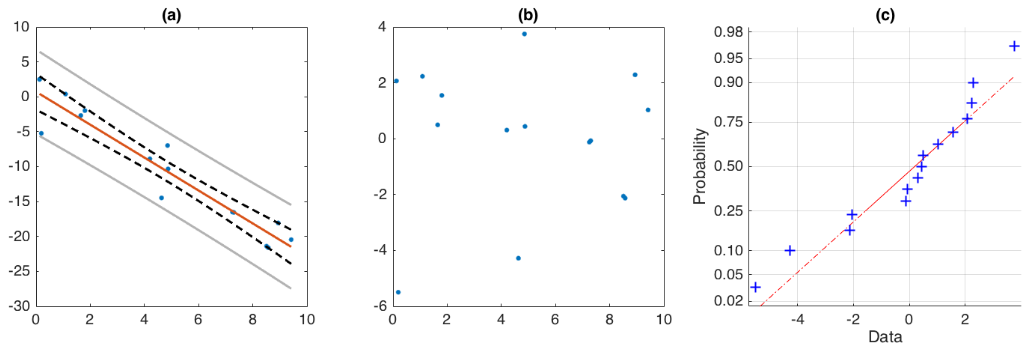

Figure 2: Resultat för Modell 2 i Uppgift 5. I (a) visas datan, skattad regressionslinje, konfidensintervall för linjens läge och prediktionsintervall. I (b) visas en plot av residualerna och i (c) visas en normalfördelningsplot av residualerna.

iii. Baserat på resultaten i Figur 2 (b) kan man dra slutsatsen att avstånd påverkar mottagen effekt.

iv. I Modell 2 antas det att y_i är log-normalfördelad.

6. Låt $X_1, \ldots, X_{10}$ vara oberoende Ge(p)-fördelade slumpvariabler och låt $Y = \sum_{i=1}^{10} Y_i$ där

$$Y_i = \begin{cases} 1 & \text{om } X_i < 3 \\ 0 & \text{annars.} \end{cases}$$

(a) I Uppgift 3(b) efterfrågades en tolkning av den negativa Binomialfördelningen i termer av Bernoulliförsök. Ange motsvarande tolkning för den geometriska fördelningen. Det finns en koppling mellan den geomtetriska fördelningen och den negativa binomialfördelningen, vilken? (1p)

(b) Antag att $p = 0.5$ och beräkna $\mathsf{P}(Y_i = 1)$ och $\mathsf{E}(Y_i)$. Bestäm också vilken fördelning Y har och beräkna $\mathsf{E}(Y)$. (3p)

(c) Använd momentmetoden för att beräkna en skattning av p baserat på en observation $Y = 7$. (3p)

(d) Använd direktmetoden för att utföra hypotestestet $H_0 : p = 0.5$, $H_1 : p > 0.5$ på nivå $\alpha = 0.05$ baserat på observationen $Y = 7$. (3p)

Lycka till!