Matematiska vetenskaper
Chalmers tekniska högskola

Tentamen: 2018-11-02 kl $08^{30} - 12^{30}$
Matematisk statistik för K, TMA074

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

Tillåtna hjälpmedel: Formelsamling och Chalmers-godkänd miniräknare.

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. Låt X beteckna antalet poäng en slumpmässigt vald student får på denna uppgift. Antag att X har en fördelning som beskrivs av följande sannolikhetsfunktion:

$$f(k) = \begin{cases} 0.1 & k = 0 \\ 0.05 & k = 1 \\ 0.1 & k = 2 \\ 0.25 & k = 3 \\ 0.2 & k = 4 \\ 0.3 & k = 5 \end{cases}$$

   Låt $\bar{X}$ beteckna medelvärdet av poängen på uppgiften bland 45 studenter som skriver tentamen, som vi antar får poäng som är oberoende och fördelade som X.

   (a) Beräkna väntevärde och varians av X. (2p)

   (b) Använd centrala gränsvärdessatsen för att approximera fördelningen för $\bar{X}$. (1p)

   (c) Använd fördelningen från (b) för att approximativt beräkna sannolikheten att medelpoängen på uppgiften överstiger 3 bland de 45 studenterna. (2p)

2. I en studie undersöks effekten av olika typer av dieter. Man mäter försökspersonernas vikter vid studiens start och igen efter två månader när personerna har följt en angiven diet. I följande tabell finns de uppmätta skillnaderna mellan försökspersonernas vikter före och efter dieten, D_i = vikt efter studien minus vikt före dieten (enhet kg).

Person	1	2	3	4	5	6	7	8	9	10
D_i	-1.43	-2.75	-1.36	-0.48	0.13	-1.56	0.08	-0.54	-0.85	2.75

   (a) Definiera medianen, M, för en kontinuerlig stokastisk variabel. (1p)

   (b) Förklara principen bakom hur ett teckentest kan användas för att undersöka om dieten minskar medianvikten hos deltagarna. (2p)

   (c) Utför teckentestet med signifikansnivå $\alpha = 0.05$ för att undersöka om dieten minskar medianvikten hos deltagarna. (2p)

3. I kursens projektet undersöktes kvicksilverhalten hos fisk fångad i Västra Götaland. Vi ska nu titta lite mer ingående på resultaten från tre sjöar nära Göteborg. I följande tabell finns mätningar av kvicksilverhalter hos gäddor i dessa tre sjöar.

Sjö	Mätningar (enhet mg/kg)
Lilla Delsjön	1.87 1.10 1.97 2.19 1.95
Stora Delsjön	1.70 0.76 0.80 0.82
Härlanda Tjärn	0.49 0.48 0.53

(a) Beräkna variansanalystabellen för datan. (4p)

(b) Skiljer sig kvicksilverhalten signifikant mellan gäddor i de tre olika sjöarna? Använd signifikansnivå $\alpha = 0.05$. (1p)

4. Stig ska byta bredbandsleverantör och behöver därför först ringa till sin nuvarande leverantör och säga upp sitt abonnemang, för att sedan ringa till den nya operatören och teckna ett nytt. Båda leverantörerna har underbemannad telefonsupport vilket gör att han måste stå i telefonkö två gånger för att utföra bytet. Låt $X \sim \text{Exp}(5)$ och $Y \sim \text{Exp}(10)$ beteckna kötiderna (i enhet minuter) för de två samtalen och antag att dessa är oberoende.

   (a) Härled den simultana täthetsfunktionen för (X, Y), $f_{X,Y}(x, y)$, och beräkna väntevärdet av den totala kötiden $Z = X + Y$. (2p)

   (b) Beräkna sannolikheten att den totala kötiden överstiger 20 minuter. (3p)

5. I ett experiment vill man undersöka hur en responsvariabel Y beror på två förklarande variabler x_1 och x_2. Man utför 15 experiment med olika värden på de förklarande variablerna för varje experiment och registrerar värdet på responsvariabeln. Man är inte säker på om x_1 verkligen påverkar responsvariabeln, och vill därför testa följande två modeller:

$$\begin{aligned} &\text{Modell 1:} \quad Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \\ &\text{Modell 2:} \quad Y_i = \beta_0 + \beta_2 x_{2i} + \varepsilon_i, \end{aligned}$$

där ε_i i båda fallen antas vara oberoende $\mathsf{N}(0, \sigma^2)$-fördelade variabler. För att kunna skatta de två modellerna beräknas följande värden:

$$\mathbf{Y}^T\mathbf{Y} = 19595, \quad \mathbf{X}^T\mathbf{Y} = \begin{pmatrix} 534.28 \\ 3394.4 \\ 1438.8 \end{pmatrix}, \quad (\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} 3.3503 & -0.4406 & -0.1982 \\ -0.4406 & 0.0754 & -0.0135 \\ -0.1982 & -0.0135 & 0.1106 \end{pmatrix},$$

$$\sum_{i=1}^{15}(x_{2i} - \bar{x}_2)^2 = 9.2437, \quad \sum_{i=1}^{15}(y_i - \bar{y})^2 = 564.66, \quad \sum_{i=1}^{15}(x_{2i} - \bar{x}_2)(y - \bar{y}) = 69.568,$$

$$\bar{x}_2 = 2.5627, \quad \bar{y} = 35.619.$$

   (a) Skatta parametrarna β_0, β_1 och β_2 i Modell 1. (2p)

   (b) Undersök om β_1 och β_2 i (a) är signifikant skilda från noll, på nivå $\alpha = 0.05$. (2p)

   (c) Skatta parametrarna β_0 och β_2 i Modell 2. (2p)

   (d) Undersök om β_2 i (c) är signifikant skild från noll, på nivå $\alpha = 0.05$. (2p)

   (e) Förklaringsgraden R^2 för Modell 1 är 0.9615. Definiera R^2 för en regressionsmodell och beräkna R^2-värdet för Modell 2. Baserat på detta och resultaten ovan, vilken modell anser du är bäst att använda för experimentet? (2p)

6. K.S. Lomax introducerade 1954 en fördelning för att modellera livslängden hos företag. Denna fördelning kallas sedan dess för Lomax-fördelningen och har bland annat också använts för att modellera trafik på internet. Täthetsfunktionen för en Lomax-fördelad slumpvariabel X är

$$f(x) = a(1 + x)^{-(a+1)},$$

där $a > 1$ är en parameter. Hos företag är det ofta så att risken för konkurs avtar med tiden. En anledning till att fördelningen sägs passa bra till att modellera livslängden hos företag att den fångar denna egenskap. Vi ska nu undersöka detta.

   (a) Härled fördelningsfunktionen för X. (2p)

(b) Antag att livslängden (i enhet 10 år) hos företag i servicebranchen kan modelleras som en Lomax-fördelad slumpvariabel med $a = 3$. Beräkna sannolikheten att ett företag går i konkurs under det första året, dvs $\mathsf{P}(X < 0.1)$. (1p)

(c) Under samma antaganden som i (b), beräkna nu sannolikheten att företaget går i konkurs under det andra året givet att det överlevde det första året. Tyder detta på att risken för konkurs avtar med tiden som företaget har funnits? (2p)

(d) Vad är den förväntade livslängden hos ett företag i servicebranchen? Alltså, beräkna väntevärdet av X då $a = 3$. (2p)

(e) Härled maximum likelihood-skattaren a^*_{ML} av a baserat på ett stickprov $x_1, \dots x_n$ observationer av oberoende Lomax-fördelade slumpvariabeler med parameter a. (3p)

Lycka till!