Tillämpad matematisk statistik LMA521 (EPI/Design/Maskin/Mekatronik-programmen) Tentamen 2018-04-06 med lösningar

Tid: 8.30-12.30. **Tentamensplats:** Lindholmen
Hjälpmedel: Kursboken **Matematisk Statistik** av Ulla Dahlbom. Formelsamlingen **Tabell- och formelsamling i matematisk statistik, försöksplanering och kvalitetsstyrning** av Håkan Blomqvist. Boken och formelsamlingen får ej innehålla extra anteckningar, men understrykningar, sticks och markeringar är tillåtna. **Chalmersgodkänd räknare.**
Examinator: Johan Tykesson
Telefonvakt/jour: Johan Tykesson, 0703182096. Till salen ca 9.30 och 11.30
Till varje uppgift skall fullständig lösning lämnas!
OBS: text på TRE sidor!
Betygsgränser: För betyg 3, 4 resp. 5 krävs minst 20, 30 resp. 40 poäng.

1. (2+4 poäng) Antag att mätvärdena 10.5, 11.8 och 9.6 kommer från en normalfördelning med väntevärde μ och standardavvikelse σ. Antag också att mätningarna är gjorda oberoende av varandra.

   (a) Beräkna ett 95% konfidensintervall för μ om $\sigma = 1$.

   (b) Beräkna ett 95% konfidensintervall för μ om σ okänt.

   Lösning:

   (a) $\bar{x} = 10.6333$. Intervallet blir

   $$10.6333 \pm 1.96 \times 1/\sqrt{3} = \boxed{10.6333 \pm 1.1316} = \boxed{[9.5017, 11.7649].}$$

   (b) $s = 1.10604$. t-tabellen, rad 2, kolumn 0.05, ger värdet 4.30. Intervallet blir

   $$10.6333 \pm 4.3 \times 1.10604/\sqrt{3} = \boxed{10.6333 \pm 2.7459} = \boxed{[7.8874, 13.3792]}$$

2. (3+3+4 poäng) Anna har fört statistik över hur långt hon går på en dag. Hon har kommit fram till att sträckan hon går på en dag kan betraktas som en kontinuerlig stokastisk variabel ξ med frekvensfunktion

   $$f(x) = \begin{cases} \frac{1}{2500}(30x^2 - 3x^3) & \text{för } 0 \leq x \leq 10 \\ 0 & \text{för övrigt} \end{cases}$$

   Enheten är kilometer.

   (a) Beräkna väntevärde och standardavvikelse för ξ.

   (b) Beräkna approximativt sannolikheten att hon går mindre än eller lika med 620 kilometer under en tidsperiod på 100 dagar. Vi antar att sträckorna hon går under olika dagar kan betraktas som oberoende.

   (c) Antag igen att vi tittar på 100 dagar. Låt η vara lika med antalet dagar av de 100 dagarna på vilka hon går längre än 5 kilometer. Beräkna approximativt $P(\eta > 65)$.

Lösning:

(a)

$$E(\xi) = \int_0^{10} xf(x)dx = ... = \boxed{6}.$$

$$\sigma = \sqrt{\int_0^{10} x^2 f(x)dx - (E(\xi))^2} = ... = \sqrt{40-36} = \boxed{2}.$$

(b) Sätt ξ_i = sträckan hon går dag nummer i, $i = 1, ..., 100$. Totala sträckan hon går under 100 dagar betecknas med T. Då är $T = \sum_{i=1}^{100} \xi_i$. Enligt centrala gränsvärdessatsen gäller att T är approximativt $N(6 \times 100, 2 \times \sqrt{100}) = N(600, 20)$-fördelad, så $(T - 600)/20$ är approx $N(0, 1)$-fördelad. Så

$$P(T \leq 620) = P(\frac{T-600}{20} \leq \frac{620-600}{20}) \approx \Phi(1) = \boxed{0.8413}.$$

(c)

$$P(\xi > 5) = \int_5^{10} f(x) = ... = \frac{11}{16}.$$

η är Bin($n = 100$, $p = 11/16$)-fördelad. Eftersom $np(1-p) = 21.5 > 10$ är η approx $N(np, \sqrt{np(1-p)}) = N(68.75, 4.63512)$. Så

$$P(\eta > 65) = 1 - P(\eta \leq 65) = 1 - P(\frac{\eta - 68.75}{4.65312} \leq \frac{65-68.75}{4.63512})$$
$$\approx 1 - \Phi(-0.809) = 1 - (1 - \Phi(0.809)) = \Phi(0.809) = \boxed{0.791.}$$

3. (6 poäng) Antag att det i urna A finns 3 röda och 4 gröna bollar, och att det i urna B finns 2 röda och 5 gröna bollar. Först väljer man en boll slumpmässigt från urna A och lägger den i urna B. Sedan väljer man en boll slumpmässigt från urna B (som nu innehåller 8 bollar), och lägger den i urna A. I det sista och tredje steget drar vi slumpässigt en boll från urna A. Vad är sannolikheten att bollen vi drar i det sista steget är grön? Vi antar hela tiden att bollarna i urnorna är väl blandade.

Lösning: Det finns 4 fall som gör att vi drar en grön i sista och tredje steget. Vi betecknar fallen med A_1, A_2, A_3 och A_4. Fallet A_1 är följande sekvens av händelser: Först flyttas en grön, sedan en grön, och en grön dras i tredje steget. A_2: grön - röd-grön. A_3: röd - grön - grön. A_4: röd - röd -grön. Sannolikheten för A_1 kan beräknas enligt följande:

$$P(A_1) = \frac{4}{7} \times \frac{6}{8} \times \frac{4}{7} = \frac{96}{392}.$$

Den första faktorn kommer sig av att vi har 4 gröna bollar av 7 bollar när vi väljer i första steget. Den andra faktorn kommer från att vi har 6 gröna av 8 i andra steget, ifall det flyttades en grön i första. Det tredje faktorn kommer från att vi har 4 gröna bollar av 7 i tredje steget ifall det flyttades gröna bollar i de två första stegen.

På liknande sätt fås att

$$P(A_2) = \frac{4}{7} \times \frac{2}{8} \times \frac{3}{7} = \frac{24}{392},$$

$$P(A_3) = \frac{3}{7} \times \frac{5}{8} \times \frac{5}{7} = \frac{75}{392},$$

$$P(A_4) = \frac{3}{7} \times \frac{3}{8} \times \frac{4}{7} = \frac{36}{392}.$$

Eftersom A_1, A_2, A_3, A_4 samtliga är disjunkta, fås

$$P(\text{grön dras i tredje steget}) = P(A_1)+P(A_2)+P(A_3)+P(A_4) = \frac{231}{392} = \boxed{\frac{33}{56}} \approx \boxed{0.5893.}$$

4. (2+2+2 poäng) Antag att ξ är en diskret stokastisk variabel. Det gäller att $P(\xi = -1) = 0.3$, $P(\xi = 1) = 0.2$, $P(\xi = 0) = 0.5$ och $P(\xi = x) = 0$ för övriga x. Vi definierar den diskreta stokastiska variabeln η genom sambandet $\eta = 1 + \xi^2$.

   (a) Beräkna väntevärdet för η.

   (b) Beräkna $P(\eta = x)$ för alla heltal x. (Det vill säga, bestäm sannolikhetsfunktionen för η.)

   (c) Beräkna den betingade sannolikheten $P(\xi = 1 \mid \eta = 2)$.

   Lösningar:

   (a)

$$E(\xi^2) = (-1)^2 \times 0.3 + 0^2 \times 0.5 + 1^2 \times 0.2 = 0.5.$$

$$E(\eta) = 1 + E(\xi^2) = \boxed{1.5}.$$

   (b) η kan anta värdena 1 och 2. Det gäller att $P(\eta = 1) = P(\xi = 0) = \boxed{0.5}$ och $P(\eta = 2) = P(\xi = 1) + P(\xi = -1) = \boxed{0.5}$ samt $P(\eta = x) = \boxed{0}$ för alla övriga x.

   (c)

$$P(\xi = 1 \mid \eta = 2) = \frac{P(\{\xi = 1\} \cap \{\eta = 2\})}{P(\eta = 2)} = \frac{P(\xi = 1)}{P(\eta = 2)} = \frac{0.2}{0.5} = \boxed{\frac{2}{5}}.$$

5. (4 poäng) För att kontrollera en kemisk tillverkningsprocess tar man med jämna mellanrum ut en provgrupp om 3 enheter och mäter pH värde. Från 10 provgrupper har man följande resultat (där $\bar{x}$ är provgruppsmedelvärdet, R är variationsbredden för provgruppen, och s är provgruppsstandardavvikelsen):

Provgrupp:	1	2	3	4	5	6	7	8	9	10
	7.0	6.9	7.0	7.2	7.0	7.0	6.9	7.0	6.9	7.1
	7.0	7.0	7.1	7.0	7.1	7.1	7.0	6.9	7.1	7.1
	7.0	7.1	6.9	6.8	6.9	7.2	7.1	7.1	7.0	7.1
$\bar{x}$	7.0	7.0	7.0	7.0	7.0	7.1	7.0	7.0	7.0	7.1
R	0.0	0.2	0.2	0.4	0.2	0.2	0.2	0.2	0.2	0.0
s	0.0	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.0

Beräkna övre och undre kontrollgränser (styrgränser) för lämpliga diagram och avgör om processen är i statistisk kontroll. Du får själv välja om du vill göra diagram som använder R-värdena eller diagram som använder s-värdena. (Egentligen brukar man mäta på åtminstone 20 provgrupper men vi har här bara 10 för att minska mängden beräkningar)

Lösning: Vi använder R-värdena i denna lösning. $\bar{\bar{x}} = 7.02$. $\bar{R} = 0.18$. $n = 3$ så konstanterna $A_2 = 1.023$, $D_3 = 0$, $D_4 = 2.575$. För $\bar{x}$-diagram blir kontrollgränserna $K_u = 7.02 - 1.023 \times 0.18 = \boxed{6.84}$ och $K_{\text{ö}} = 7.02 + 1.023 \times 0.18 = \boxed{7.20}$. För R-diagram blir $K_u = \boxed{0}$ och $K_{\text{ö}} = 2.575 \times 0.18 = \boxed{0.4653}$. Alla $\bar{x}$ värden ligger inom gränserna för $\bar{x}$-diagrammet, och alla R-värden ligger inom gränserna för R-diagrammet. Därför är processen i statistisk kontroll.

6. (2+2+1 poäng) En djurpark skall köpa in ett parti på 20 nya flodhästar. Naturligtvis bör alla 20 flodhästar veterinärundersökas innan de köps in, men djurparken beslutar sig för att chansa. De använder en enkel provtagningsplan där man undersöker ett urval på 3 flodhästar. Om alla 3 är friska accepteras partiet, annars avvisas det. Om partiet avvisas så gör man en allkontroll av partiet, dvs samtliga 20 flodhästar undersöks. Antag att 3 av de 20 flodhästarna är sjuka.

   (a) Beräkna sannolikheten att flodhäst-partiet accepteras av den enkla provtagningsplanen.

   (b) Beräkna väntevärdet för antalet undersökta flodhästar (dvs, beräkna ATI).

   (c) Låt d beteckna antalet sjuka flodhästar i urvalet. Beräkna standardavvikelsen för d.

   Lösningar: Sätt d = antalet sjuka flodhästar i urvalet. d är $Hyp(N = 20, n = 3, p = 3/20)$.

   (a)
   $$P(d = 0) = \frac{17 \times 16 \times 15}{20 \times 19 \times 18} = \boxed{0.5965.}$$

   (b)
   $$ATI = 3 \times 0.5965 + 20 \times (1 - 0.5965) = \boxed{9.8595}.$$

   (c) Eftersom d är hypergeometriskt fördelad med parametrar enligt ovan fås att standardavvikelsen för d ges av
   $$\sqrt{np(1-p)\frac{N-n}{N-1}} = \boxed{0.5850}.$$

7. (6 poäng) Följande gäller för en viss typ av datachips. De flesta datachipsen genomgår en kvalitetskontroll innan de går till försäljning, och repareras om de är defekta. Ibland misslyckas reparationen. Dessutom finns det datachips som inte kontrolleras alls.

   Antag att sannolikheten att ett datachips genomgår kvalitetskontroll är 0.99. Sannolikheten att ett datachips som genomgått kvalitetskontroll är defekt är 0.02. Sannolikheten att ett datachips som inte genomgått kvalitetskontroll är defekt är 0.12.

Antag att du köper ett datachips, och ser att det är defekt. Beräkna den betingade sannolikheten att datachipset genomgått kvalitetskontroll givet detta.

Lösning: Sätt $D = \{\text{chips defekt}\}$ och $K = \{\text{chips kontrollerat}\}$. Vi vet att $P(K) = 0.99$, $P(D|K) = 0.02$ och $P(D|K^c) = 0.12$. Vi får att

$$P(D) = P(D|K)P(K)+P(D|K^c)P(K^c) = 0.02\times 0.99+0.12\times 0.01 = 0.021.$$

Enligt Bayes sats blir

$$P(K|D) = \frac{P(D|K)P(K)}{P(D)} = \frac{0.02 \times 0.99}{0.021} = \boxed{0.9429}.$$

8. (3+4 poäng) Man genomförde ett fullständigt faktorförsök för att undersöka hur de 3 faktorerna A, B och C påverkade en speciell situation. Man fick följande resultat från de åtta försöken.:

Nr.	A	B	C	Resultat y
1	-	-	-	41
2	+	-	-	51
3	-	+	-	42
4	+	+	-	51
5	-	-	+	38
6	+	-	+	55
7	-	+	+	40
8	+	+	+	51

(a) Beräkna huvudeffekten l_A, två-faktorsamspelet l_{AB}, och tre-faktorsamspelet l_{ABC}.

(b) Antag att man också var intresserad av faktorerna D och E. Man har bara råd att göra 8 försök, så man får göra ett reducerat faktorförsök. Antag att man väljer teckenkolumner för A, B och C precis som ovan. Det finns sedan olika sätt att välja teckenkolumner för D och E. Ingenjörerna Benedikt och Beata har fått i uppdrag att lägga upp det reducerade faktorförsöket med kravet att A inte sammanblandas med BC. Benedikt väljer generatorerna $D = AB$ och $E = AC$, medan Beata väljer generatorerna $D = ABC$ och $E = AB$. Kommer Benedikts reducerade faktorförsök att uppfylla villkoret? Kommer Beatas reducerade faktorförsök att uppfylla villkoret? Beräkna också upplösningarna för de bägge planerna.

Lösningar:

(a) $\boxed{l_A = 11.75}$, $\boxed{l_{AB} = -1.75}$, $\boxed{l_{ABC} = -1.25}$

(b) Benedikt: $I_1 = ABD$, $I_2 = ACE$, $I_3 = ABDACE = BCDE$. Alias till A blir BD, CE och $ABCDE$. Uppfyller således kravet, och upplösningen blir III.

Beata: $I_1 = ABCD$, $I_2 = ABE$, $I_3 = ABCDABE = CDE$. Alias till A blir BCD, BE och $ACDE$. Uppfyller således kravet, och upplösningen blir III.

Lycka till!