Lösningsförslag till Tillämpad matematisk statistik LMA521, Tentamen 20180313

Betygsgränser: för betyg 3 krävs minst 20 poäng, för betyg 4 krävs minst 30 poäng, för betyg 5 krävs minst 40 poäng.

1. Vid en kvalitetskontroll av $N = 1500$ kantträdsdäck undersöktes $n = 100$. Antal felaktiga däck bland de 1500 däcken var 35, så att $p = \frac{35}{1500} \ldots$ Sätt $x = 3$

   (a) Använd hypergeometrisk fördelning. Ett exakt uttryck för sannolikheten att exakt 3 av de 100 utvalda däcken är felaktiga är

   $$\frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}} = \frac{\binom{35}{3}\binom{1465}{97}}{\binom{1500}{100}}$$

   3p

   (b) Sannolikheten i (a) approximativt med lämplig binomialfördelning är ($n/N = 100/1500 < 0.1$)

   $$\binom{n}{x}p^x(1-p)^{n-x} = \binom{100}{3}p^3(1-p)^{97} = 0.207988\ldots$$

   Svar: Sökt sannolikhet är 0.21. 3p

2. Följande funktion $f(x)$ är en frekvensfunktion för en stokastisk variabel ξ.

   $$f(x) = \begin{cases} C\,x^3(1-x), & \text{om } 0 \le x \le 1 \\ \\ 0, & \text{för övrigt} \end{cases}$$

   är given.

   (a) Konstanten C:

   $$1 = C\int_0^1 (x^3 - x^4)dx = C\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 = C \cdot \frac{1}{20}.$$

   Svar: $C = 20$. 2p

   (b) Väntervärdet

   $$E(\xi) = 20\int_0^1 (x^4 - x^5)dx = 20\left[\frac{x^5}{5} - \frac{x^6}{6}\right]_0^1 = 20 \cdot \frac{1}{30} = \frac{2}{3}.$$

   2p

3. Givet är fem oberoende mätningar som gav värdena 6.5, 6.3, 4.2, 5.5, 6.0 av en normalfördelad stokastisk variabel. Medelvärdet är $\overline{x} = 5.7$. Ett (symmetriskt) 95%:s konfidensintervall för μ

   (a) om $\sigma = 0.9$

   $$[\overline{x} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x}] + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}] = [5.7 - 1.96 \cdot \frac{0.9}{\sqrt{5}}, 5.7 + 1.96 \cdot \frac{0.9}{\sqrt{5}}] = [4.9.6.5].$$

   2p

   (b) om σ okänd ersätts σ med $s = 0.919239$ och $\lambda_{\alpha/2}$ ersätts av $t_{\alpha/2}(4) = 2.78$.

   $$[\overline{x} - t_{\alpha/2}(4)\frac{s}{\sqrt{n}}, \overline{x}] + _{\alpha/2}(4)\frac{s}{\sqrt{n}}] = [5.7 - 2.78 \cdot \frac{0.919\ldots}{\sqrt{5}}, 5.7 + 2.78 \cdot \frac{0.919\ldots}{\sqrt{5}}] = [4.5.6.9].$$

   4p

4. Givet är

$$P(K) = 0.50, \quad P(K|W^c) = 0.20, \quad P(W|K) = 0.90$$

(a) Beräkna sannolikheten att ett däck är korrekt ochr av märket W...

$$P(W \cap K) = P(W|K) \cdot P(K) = 0.45.$$

2p

(b) Beräkna sannolikheten $P(K|W)$...

$$P(K|W) = \frac{P(K \cap W)}{P(W)}$$

Täljaren är beräknad i (a) och är 0.45. Vi måste beräkna $P(W)$.
Enligt Bayes sats är

$$P(K|W^c) = \frac{P(W^c|K)P(K)}{P(W^c)}.$$

Eftersom $P(W^c|K) = 1 - P(W|K) = 0.1$ får vi att

$$0.2 = \frac{0.1 \cdot 0.5}{P(W^c)},$$

vilket get att $P(W^c) = 0.25$ så att $P(W) = 0.75$. Slutligen fås att $P(K|W) = 0.45/0.75 = 0.6$. 5p

5. En registreringsskylt har sex tecken , först tre bokstäver (valda bland 23) och sedan tre siffror (valda bland 0,1,2,3,4,5,6,7,8,9).

(a) Antal registreringsnummer:

$$23^3 \cdot 10^3 = 1\,2167\,000.$$

1p

(b) Hur många registreringsnummer har minst två lika tecken (d.v.s minst två lika bokstäver eller minst två lika siffror? Komplementhändelsen är att alla bokstäver olika och alla siffror olika. Detta antal är

$$10 \cdot 9 \cdot 8 \cdot 23 \cdot 22 \cdot 21 = 7650720.$$

Sökt antal är

$$1\,2167\,000 - 7650720 = 4\,516\,280 (\approx 4.5 \cdot 10^6).$$

3p

(c) Hur många registreringsnummer har minst två lika tecken (d.v.s minst två lika bokstäver och minst två lika siffror?

$$(23^3 - 23 \cdot 22 \cdot 21)(10^3 - 10 \cdot 9 \cdot 8) = 431\,480$$

3p

6. (8 poäng)

(a) Så de tre kolumnerna måste tillsammans ge alla möjliga kombinationer av faktorerna A,B och C på låg resp. hög nivå på endast 8 rader. Detta kan vi åstadkomma genom att t.ex. sätta dem såsom i tabell 1.

(b) Trefaktorsamspelet kan räknas ut genom att för varje rad i de andra kolumnerna multiplicera dem med varandra (Tänk på att + egentligen betyder +1 och - egentligen betyder -1). Alltså blir första raden $(-1) \cdot (-1) \cdot (-1) = -1$. Det blir alltså ett + på första raden av ABC. Fortsätter man vidare får man kolumnen ABC i tabell 1.

(c) Vi vill alltså konstruera en 2^{4-1}-plan. Nu gäller det att välja generatorn på ett smart sätt så att inga huvudeffekter får ett alias som också är en huvudeffekt eller tvåfaktor-samspelseffekt.

Eftersom vi bara reducerar en nivå så kommer vi bara ha en definierande relation. Vi ser nu att om ordet i den definierande relationen har 4 bokstäver så kommer huvudeffekter att sammanblandas antingen med trefaktor-samspel eller fyrafaktor-samspel. Detta beror ju på att en bokstav multiplicerat med fyra bokstäver inte kan ta bort mer än maximalt en av bokstäverna. Därför väljer vi generatorn $D = ABC$ som ger oss $I = ABCD$.

(d) Upplösningen blir IV eftersom vi har fyra bokstäver i ordet i vår definierande relation.

(e) Aliasen till huvudeffekterna hittar vi genom att multiplicera dem med ordet i den definierande relationen. T.ex. så blir aliaset till A $A \cdot ABCD = BCD$. De resterande sammanblandningsmönstren hittar du i tabell 2.

A	B	C	ABC
-	-	-	-
+	-	-	+
-	+	-	+
+	+	-	-
-	-	+	+
+	-	+	-
-	+	+	-
+	+	+	+

Tabell 1

Effekt	Effekt $\cdot I$	Alias
A	$A = A \cdot ABCD$	BCD
B	$B = B \cdot ABCD$	ACD
C	$C = C \cdot ABCD$	ABD
D	$D = D \cdot ABCD$	ABC

Tabell 2: Sammanblandningsmönster för huvudeffekter.

7. (6 poäng)

(a) Vi har att $p_1 = 2\%$, $\alpha = 5\%$, $p_2 = 9\%$ och $\beta = 10\%$. Vi tittar i tabellerna för dubbelprovtagningsplan då de matchar de producent- och konsument-riskerna vi är beredda att ta. Vi har en plan för $n_1 = n_2$ och en för $n_2 = 2n_1$ och vi tittar i båda. Först måste vi räkna ut kvoten $\frac{p_2}{p_1} = \frac{9}{2} = 4.5$. Inget värde i tabellerna är exakt 4.5 så vi vill välja närmaste värdet som är lite mindre än 4.5. Anledningen till att vi väljer ett litet mindre värde är att vi då kommer få en konsumentrisk som är mindre än vad vi kräver. Vi är alltså på den säkra sidan. Det innebär dock att vår provplan kommer ha ett lite större n än vad vi egentligen skulle behöva. I tabellen för $n_2 = 2n_1$ så är provtagningsplan 6 lämplig då den har en kvot på 4.31 (närmaste i den andra tabellen hade varit 4.25 vilket är längre ifrån 4.5).

Vi säger alltså att $c_1 = 0$, $c_2 = 4$ och $n_1p_1 = 0.68$. Från detta vet vi att $n_1 = \frac{0.68}{0.02} = 34$, $n_2 = 68$ och $r_1 = r_2 = c_2 + 1 = 5$.

(b) De frågar alltså efter $ASN(2\%)$,

$$ASN(2\%) = n_1\mathbb{P}(\text{beslut i urval 1}) + (n_1 + n_2)\mathbb{P}(\text{beslut i urval 2}).$$

Låt oss modellera antal defekta funna i första urvalet med slumpvariabeln $\xi \sim Bin(34, 2\%)$. Här använde vi oss avbinomialapproximation eftersom $\frac{34+68}{10^4} < 10\%$. Vi kan nu räkna ut sannolikheten att bestämma oss i urval 2 som

$$\mathbb{P}(\text{beslut i urval 2}) = \mathbb{P}(0 < \xi < 5) = \mathbb{P}(1 \leq \xi \leq 4)$$

$$\mathbb{P}(\xi = 1) = \binom{34}{1} 0.02^1 \cdot 0.98^{33} = 34.91\%$$
$$\mathbb{P}(\xi = 2) = \binom{34}{2} 0.02^2 \cdot 0.98^{32} = 11.76\%$$
$$\mathbb{P}(\xi = 3) = \binom{34}{3} 0.02^3 \cdot 0.98^{31} = 2.60\%$$
$$\mathbb{P}(\xi = 4) = \binom{34}{4} 0.02^4 \cdot 0.98^{30} = 0.41\%$$

Därför är $\mathbb{P}(1 \leq \xi \leq 4) = 34.91\% + 11.76\% + 2.60\% + 0.41\% = 49.68\%$. ASN kan då räknas ut som $ASN(2\%) = 34 \cdot 0.5032 + 102 \cdot 0.4968 = 67.78$. Så i genomsnitt så kommer ungefär 68 helikoptrar kontrolleras i varje parti om felkvoten verkligen skulle vara 2%.

(c) Standardavvikelsen är ju alltid roten ur variansen. Variansen kan vi räkna ut med hjälp av sannolikheterna från b-uppgiften.

$$Var(\text{antal kontrollerade}) = \mathbb{E}\left[\text{antal kontrollerade}^2\right] - \mathbb{E}\left[\text{antal kontrollerade}\right]^2.$$

Därför är

$$Var(\text{antal kontrollerade}) = 34^2 \cdot 0.5032 + 102^2 \cdot 0.4968 - 67.78^2 = 1156.278.$$

Standardavvikelsen blir då $\sqrt{1156.278} \approx 34$.

8. (6 poäng)

OBS: I lösningen räknar vi med att parametrarna μ och σ är kända, som står angivet i texten. Har man räknat med okända parametrar kan detta också ge poäng.

(a) Vi känner till det sanna μ och σ. Därför drar vi centrallinjen i μ och över och undre styrgränser som

$$S_{\text{ö}} = \mu + 3\frac{\sigma}{\sqrt{n}} = 13 + 3 * \frac{3}{\sqrt{7}} = 16.4,$$
$$S_{\text{u}} = 13 - 3\frac{3}{\sqrt{7}} = 9.6.$$

För s-diagrammet så sätter vi över och undre gränser med hjälp av σ och två konstanter som vi kan hitta i tabellen för kontrolldiagram, $S_{\text{ö}} = B_6\sigma$ och $S_{\text{u}} = B_5\sigma$. Eftersom vi har 7 mätningar i varje provgrupp så tittar vi på raden som motsvarar provgruppstorlek 7. Där ser vi att $B_5 = 0.113$ och $B_6 = 1.806$. Alltså får vi

$$S_{\text{ö}} = 1.806 \cdot 3 = 5.418$$
$$S_{\text{u}} = 0.113 \cdot 3 = 0.339.$$

Detta resulterar i diagram såsom i figur 1.

(b) Målvärdet $M = 14$, $T_{\text{ö}} = 14 + 4 = 18$ och $T_{\text{u}} = 14 - 4 = 10$. Vi kan nu räkna ut duglighetsindexet som

$$C_p = \frac{T_{\text{ö}} - T_{\text{u}}}{6\sigma} = \frac{8}{6 \cdot 3} = 0.44.$$

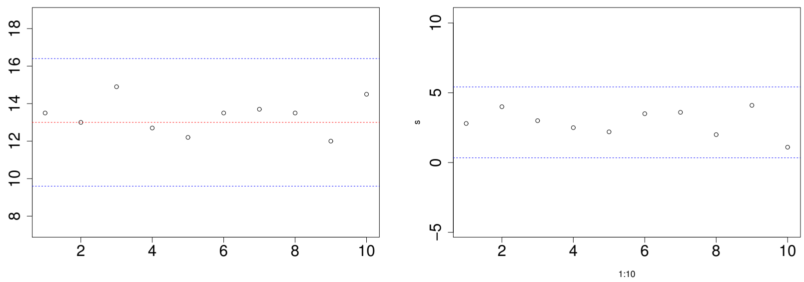
För att räkna ut det korrigerade duglighetsindexet måste vi titta på centreringen av processen,

$$CM = 2\frac{|M - \mu|}{T_{\text{ö}} - T_{\text{u}}} = \frac{1}{4}.$$

Korrigerat duglighetsindex blir då

$$C_{pk} = C_p(1 - CM) = 0.44 \cdot 0.75 = 0.33.$$

Eftersom $C_p < 1.33$ så är vår spridning i processen för stor för våra kvalitetskrav. Processen är alltså inte acceptabel. Eftersom $C_{pk} < C_p < 1.33$ så kan man önska även en bättre centrering av processen.



(a) Medelvärdesdiagram

(b) s-diagram

Figur 1