Lösningsförslag till tentamen 20190610 14.00-18.00, LMA522

1.
$$E(\xi) = \int_0^1 xf(x)dx = \int_0^1 6x^2(1-x)dx = \ldots = \frac{1}{2}. \tag{1}$$
$$E(\xi^2) = \int_0^1 x^2 f(x)dx = \int_0^1 6x^3(1-x)dx = \ldots = \frac{3}{10}. \tag{2}$$
$$\sigma = \sqrt{E(\xi^2) - E(\xi)^2} = \sqrt{3/10 - 1/4} = \frac{1}{2\sqrt{5}} \approx 0.22. \tag{3}$$

3p

2. Bilar av ett visst märke råkar ut för fel $P(A) = 0.15$ och fel $P(B) = 0.12$ oberoende av varandra. Beräkna följande sannolikheter...
$P(A \cap B) = \{\text{ober.}\} = P(A) \cdot P(B) = 0.018$.

   (a) Sannolikheten för minst ett fel är $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.15 + 0.12 - 0.018 = 0.252$.

   (b) Sannolikheten för inget fel är $P((A \cup B)^c) = 1 - P(A \cup B) = 0.748$.

   (c) Sannolikheten för exakt ett fel är $P(A) + P(B) - 2P(A \cap B) = 0.234$.

   (d) $P(A|B) = P(A) = 0.15$ p.g.a. oberonde.

4p

3. Tiden för ny asfaltsläggning av Götaälvbron beräknas ta tiden $\xi \in \mathrm{N}(3.0, 0.42)$ veckor och tiden för att måla väglinjer beräknas ta tiden $\zeta \in \mathrm{N}(1.0, 0.40)$ veckor. Dessa två tidsåtgångar är oberoende varandra.

   (a) Sannolikheten att asfaltsläggningen tar högst 3.5 veckor är $\Phi\left(\frac{3.5-3.0}{0.42}\right) = \Phi(1.19) = 0.88$. 2p

   (b) $\xi + \zeta \in \mathrm{N}(7.0, 0.58)$. Sannolikheten att den totala tiden, d.v.s. tiden för asfaltläggning och vägmålning, tar minst 4.5 veckor är

$$1 - \Phi\left(\frac{4.5-4.0}{0.58}\right) 1 - \Phi(0.86) = 0.19$$

3p

4. Antal inkommande paket till en postterminal per dag är Poissonfördelat $\mathrm{Po}(\lambda)$ med väntevärdet $\lambda = 30.0$. Använd lämpliga approximationer nedan.

   (a) Sannolikheten att det under en dag kommer in exakt 30 paket är

$$P(\xi = 30) = \frac{e^{-30} \cdot 30^{30}}{30!} = 0.073.$$

1p

   (b) Sannolikheten att det under en dag kommer in fler än 30 paket är, med normalfördelning, approximativt

$$P(\xi \geq 30.0) = 1 - \Phi\left(\frac{30-30}{\sqrt{30}}\right) = 0.5.$$

   (Utan approximation 0.451648....) 2p

(c) Sannolikheten att det under en dag har kommit högst 40 paket givet att det har kommit minst 30 paket är

$$p = \frac{P(30 \le \xi < 40)}{P(\xi \ge 30)} = \frac{\Phi\left(\frac{40.0}{\sqrt{30}}\right) - 0.5}{0.5} = 0.93$$

3p

Räknehjälp: Medelvärdet är 34.5 och standardavvikelsen är 1.5.

5. (a) För ett tvåsidigt 95%:s konfidensintervall för väntevärdet μ behövs kvantilen $t_{0.025,4} = 2.78$.

$$\left[34.5 - \frac{1.5 \cdot 2.78}{\sqrt{5}}, 34.5 + \frac{1.5 \cdot 2.78}{\sqrt{5}}\right] = [32.6,\ 36.4].$$

2p

(b) För ett uppåt begränsat 95%:s konfidensintervall för standardavvikelsen σ med $5 - 1 = 4$ frihetsgrader, behövs kvantilen $\chi^2_{0.05,4} = 0.71$. Intervallet är

$$\left[0, \sqrt{\frac{(5-1)}{\chi^2_{0.05,4}}} \cdot 1.5\right] = [0, 3.6\,]$$

4p

6. (1+2+2+1 poäng)

(a) Beräkna...

$$P(A) = \frac{P(A \cap B)}{P(B|A)} = \frac{0.4}{0.8} = 0.5.$$

(b) Beräkna...

$$P(B) = 1 - P(B^c) = 1 - \frac{P(A \cap B^c)}{P(A|B^c)} = 1 - \frac{P(A) - P(A \cap B)}{P(A|B^c)} = 1 - \frac{0.5 - 0.4}{0.25} = 0.6.$$

(c) För en tredje händelse C gäller att $P(A^c|C) = P(C|A^c)$. Beräkna $P(C)$:

$$\begin{cases} P(C|A^c) = \dfrac{P(C \cap A^c)}{P(A^c)} \\ \\ P(A^c|C) = \dfrac{P(C \cap A^c)}{P(C)} \end{cases} \Longrightarrow 1 = \frac{P(C)}{1 - P(A)} \Longleftrightarrow P(C) = 1 - P(A) = 0.5\,.$$

(d) Är A och C oberoende? Argumentering:

$$P(A \cap C^c) = 0.16 \text{ och } P(C^c) \cdot P(A) = 0.25$$

alltså beroende.

7. (2 + 2 + 2 + 2 poäng)

En mystisk alkemist har sitt laboratorium uppe i ett mörkt tornrum i ett slott på den skotska landsbygden. Där uppe har hon upptäckt formeln för att göra guld av gamla petflaskor och cigarettfimpar.

I hennes formel är renheten guld responsvariabeln, Y (som mäts i karat). Y kan anses approximativt normalfördelad med varians σ^2 och ett väntevärde som beror på faktorernas nivåer. Faktorerna är mängden petflaskor (A), mängden cigarettfimpar (B) samt mängden av en hemlig vätska (C). Formeln för väntevärdet av responsvariablen givet faktorernas inställningar ges av

$$\beta_0 + \beta_A x_A + \beta_C x_C + \beta_{BC} x_B x_C. \tag{4}$$

Här motsvarar t.ex. x_c ett värde (antingen $+1$ eller -1) beroende på om mängden av vätska C är på låg eller hög nivå.

(a) Givet formeln i Ekvation (4), beskriv vilka effekter som kan existera (alltså vilka effekter som vi vet inte bara är noll).

Lösning

De existerande effekterna är A, C och BC.

(b) För att ta reda på effekternas storlek så utförde alkemisten ett reducerat faktorförsök med två nivåer, 2^{3-1}, såsom i tabell 1.

Man kan tänka sig faktorförsöket som ett vanligt 2^2-faktorförsök för faktorerna A och C där man sedan lagt till faktor B så att den alltid har samma nivå ($+1$ eller -1) som A i varje mätning. Vilken generator och upplösning för den reducerade försöksplanen ger detta?

Nr.	A	B	C	y	s^2
1	+	+	+	$[15.16, 17.71]$	3.26
2	-	-	+	$[3.38, 1.40]$	1.98
3	+	+	-	$[10.75, 11.53]$	0.31
4	-	-	-	$[18.07, 17.37]$	0.24

Table 1: Faktorförsök med 2 replikat i varje mätgrupp.

Lösning

Generatorn $A = B$. Detta ger order $I = AB$ och därför upplösningen II, alltså en 2^{3-1}_{II}-plan.

(c) Skriv upp sammanblandningsmönstret (alltså vilka effekter som är varandras alias)

Lösning

Ordet är $I = AB$. Multiplicerar vi en effekt med ordet får vi dess alias. Detta ger sammanblandningsmönstret som i tabell 2.

(d) Med hjälp av din kunskap om vilka effekter som existerar, skatta effekterna A, C och BC samt medelvärdet M.

M	AB
A	B
C	ABC
AC	BC

Table 2: Sammanblandningsmönster. Raderna motsvarar olika grupper av beblandade effekter. Kolumnerna motsvarar effekter som är alias till varandra.

Lösning

$$
\begin{aligned}
M &= \frac{15.16+17.71+18.07+17.37+3.38+1.40+10.75+11.53}{8} = 11.92\\
l_A &= \frac{15.16+17.71+10.75+11.53}{4} - \frac{3.38+1.40+18.07+17.37}{4} = 3.73\\
l_C &= \frac{15.16+17.71+3.38+1.40}{4} - \frac{10.75+11.53+18.07+17.37}{4} = -5.02\\
l_{BC} &= \frac{15.16+17.71+18.07+17.37}{4} - \frac{3.38+1.40+10.75+11.53}{4} = 10.31.
\end{aligned}
$$

8. (2 + 4 poäng)

   Förutom tornet, som bebos av alkemisten, så används slottet i föregående uppgift som hotell och restaurang åt besökande turister. Hotellet är speciellt känt för det utsökta vin som de serverar till middagarna. Hemligheten till att deras vin håller så hög kvalitet är att hovmästaren är utbildad i statistisk kvalitetsstyrning. Därför utför han acceptanskontroll på varje parti med vinflaskor som de köper in.

   De använder sig av en dubbelprovtagningsplan med producentrisk, $\alpha = 5\%$, konsumentrisk, $\beta = 10\%$ och ett andra urval som är dubbelt så stort som det första urvalet, $n_2 = 2n_1$.

   (a) Hovmästaren, som inte spottar i glaset, utför själv kontrollen på samtliga partier. För den dubbla provtagningsplanen han valt så är $p_1 = 2\%, c_1 = 1, c_2 = 3, r_1 = r_2 = 4$.

   Med hjälp av den information du har och en lämplig tabell, ta reda på värdena hos n_1, n_2 och p_2.

   ### Lösning

   Provtagningsgrupp nummer 3 i tabellen för $n_2 = 2n_1$ stämmer in på den givna provtagningsplanen. Då ser man att $\frac{p_2}{p_1} = 6.48$ vilket ger $p_2 = 12.06\%$. Man kan även från tabellen se att $n_1 p_1 = 0.6$ vilket ger $n_1 = 30$. Således är $n_2 = 60$.

   (b) Vad är acceptanssannolikheten för en egentlig felkvot på $p = 10\%$?

Lösning

Låt ε_1 vara antalet defekta flaskor i första urvalet. Låt på samma sätt ε_2 vara antalet defekta flaskor i andra urvalet.

Enligt antagandet om $0.1N > n_1 + n_2$ så kan vi använda oss av binomialapproximationen, alltså

$$\varepsilon_1 \sim Bin(n_1, p)$$
$$\varepsilon_2 \sim Bin(n_2, p).$$

Acceptanssannolikheten kan nu beräknas genom följande formel,

$$L(p = 10\%) = \mathbb{P}\left(\varepsilon \leq 1\right) + \sum_{i=2}^{3} \mathbb{P}\left(\varepsilon_2 \leq 3 - i\right) \mathbb{P}\left(\varepsilon_1 = i\right).$$

För att beräkna sanolikheten behöver vi först räkna ut alla delsannolikheter,

$$\mathbb{P}\left(\varepsilon_1 = 0\right) = 0.9^{30} = 4.24\%$$
$$\mathbb{P}\left(\varepsilon_1 = 1\right) = \binom{30}{1} 0.1 \cdot 0.9^{29} = 14.13\%$$
$$\mathbb{P}\left(\varepsilon_1 = 2\right) = \binom{30}{2} 0.1^2 \cdot 0.9^{28} = 22.77\%$$
$$\mathbb{P}\left(\varepsilon_1 = 3\right) = \binom{30}{3} 0.1^3 \cdot 0.9^{27} = 23.61\%$$
$$\mathbb{P}\left(\varepsilon_2 = 0\right) = 0.9^{60} = 0.18\%$$
$$\mathbb{P}\left(\varepsilon_2 = 1\right) = \binom{60}{1} 0.1 \cdot 0.9^{59} = 1.20\%.$$

Genom att lägga in dessa sannolikheter i formeln får vi slutligen svaret,

$$L(p = 10\%) = (4.24\% + 14.13\%) + (0.18\% + 1.20\%)\,(22.77\%) + (0.18\%)\,(23.61\%) = 18.73\%.$$

Det är alltså lite mindre än 20% sannolikhet att acceptera ett parti vin där 10% flaskorna innehåller dåligt vin.

9. (3 + 1 + 1 + 1 poäng)

Det skotska slottshotellet i föregående uppgifter har problem med spöken. Under en period var det så illa att hotellgäster krävde pengarna tillbaka men efter åtskilliga behandlingar av både byns äldre och traktens mest slipade dammsugarförsäljare så är nu spökerierna nere på en acceptabel nivå.

Hotellet utför styrande kontroll för att garantera att spökerierna inte förvärras igen. Detta görs genom att varje natt avlyssna 6 stycken slumpvist utvalda sovrum. Kvalitetsindikatorn, x_{ij} för natt i och rum j, är antal minuter som det hörs ljud av tunga suckningar, rasslande kedjor eller klagande kvinnosång.

Medelvärdet och variationsbredden hos tiden av inspelade spökljud under 20 nätter kan ses i tabell 3.

i	1	2	3	4	5	6	7	8	9	10
$\bar{x}_i$	6.27	6.20	7.22	4.92	5.22	5.20	5.98	6.47	8.37	7.28
R_i	5.28	8.19	5.83	9.33	3.49	8.47	8.54	5.41	8.45	5.07
i	11	12	13	14	15	16	17	18	19	20
$\bar{x}_i$	7.05	6.26	5.25	8.14	8.11	6.90	7.16	9.30	10.61	12.49
R_i	7.36	6.50	7.99	7.72	10.21	7.73	8.50	7.11	7.23	3.54

Table 3: Medelvärdet och variationsbredden hos antal minuter inspelade spökljud. Kolumnerna motsvarar olika nätter och under varje natt har 5 stycken sovrum avlyssnats.

(a) Med hjälp av tabell 3 och passande tabell från formelsamlingen, räkna ut styrgränserna för både medelvärdesdiagram och passande spridningsdiagram.

Lösning

Eftersom $n = 6$ så hittar vi i lämplig tabell i formelsamlingen, $A_2 = 0.483, D_3 = 0$ och $D_4 = 2.004$.

$$\bar{\bar{x}} = \frac{1}{20}\sum_{i=1}^{20} \bar{x}_i \approx 7.22$$

$$\bar{R} = \frac{1}{20}\sum_{i=1}^{20} R_i \approx 7.10$$

För medelvärdesdiagrammet:

$$S_{\text{ö}} = \bar{\bar{x}} + A_2\bar{R} = 7.22 + 0.483 \cdot 7.10 = 10.65$$
$$S_{\text{u}} = \bar{\bar{x}} - A_2\bar{R} = 7.22 - 0.483 \cdot 7.10 = 3.79.$$

För R-diagrammet:

$$S_{\text{ö}} = D_4\bar{R} = 2.004 \cdot 7.10 = 14.22$$
$$S_{\text{u}} = D_3\bar{R} = 0.$$

(b) Är spökerierna under statistisk kontroll? Motivera ditt svar med hjälp av de uträknade styrgränserna.

Lösning

Mätningarna under natt 20 har ett medelvärde som ligger över den övre gränsen hos medelvärdesdiagrammet. Alltså, processen är inte under statistisk kontroll. Det verkar som att spökerierna ökar.

(c) Förklara med ord vad ARL är.

Lösning

ARl står för “average run length”. Det är den förväntade tiden (i antal mätningstillfällen) till ett larm uppkommer. ARL beror på fördelningen av kvalitetsindikatorn jämfört med styrgränserna.

(d) Vad är fördelen respektive nackdelen med att ha ett stort ARL.

Lösning

Ett stort ARL är positivt för en process som är under statistisk kontroll. Det innebär att den förväntade tiden till ett falsklarm är lång. Detta minskar kostnaderna associerade med att söka efter en förändring (som faktiskt inte har skett).

Om processen inte är under statistisk så vill man upptäcka det så fort som möjligt. Därför är det dåligt om ARL är stor för en process som inte är under statistisk kontroll. Det innebär att man, i genomsnitt, kommer få vänta länge innan den styrande kontrollen upptäcker förändringen.