Lösningsförslag till Tillämpad matematisk statistik LMA521
Tentamen 20190115

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

---

1. (2+2+3 poäng) För händelserna A och B gäller att

$$P(A) = 0.5, \quad P(B) = 0.4, \quad P(A \cap B) = 0.3.$$

Beräkna...

(a)

$$P(A|B^c) = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{0.5 - 0.3}{1 - 0.4} = 0.33\ldots$$

(b)

$$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)] = 0.4$$

(c)

$$\begin{aligned} P((A \cup B) \cap (A \cap B)^c) &= P((A \cup B) \setminus (A \cap B)) = \{\text{även}\} = \\ &= P(A \setminus B) + P(B \setminus A) = P(A) + P(B) - 2P(A \cap B) = 0.3. \end{aligned}$$

2. (1+3 poäng) I de lilla municipalsamhället Stuvesberg är telefonnumren femsiffriga, från 20 000 till 99 999.

   (a) Antal telefonnummer, som är möjliga (finns) är 80 000.

   (b) Man behåller alla telefonnummer som är möjliga i a) men tar bort de som börjar med siffran 7, d.v.s. $80\,000 - 10\,10\,000 = 70\,000$. Därefter inför man alla sexsiffriga telefonnummer, som börjar med 7, d.v.s. man får 100 000 nya telefonnummer.

   Antal möjliga telefonnummer finns det totalt efter dessa två ändringar är alltså 170 000.

3. (2+2+3 poäng) Tiden för Börje att lägga ett golv i ett rum (på 60m^2), är normalfördelad med väntevärde 1.5 och standardavvikelse 0.40, dvs $\xi_1 \in \text{N}(1.5, 0.40)$ (enhet timmar). Efter att golvet är lagt, tapetseras rummet av Lasse och tar en tid, som är $\xi_2 \in \text{N}(2.5, 0.41)$ (enhet timmar).

   (a) Sannolikheten golvläggningen tar mindre än 2.0 h är

$$\Phi\left(\frac{2.0 - 1.5}{0.40}\right) = 0.89.$$

   (b) För sannolikheten att den totala tiden (golvläggning+tapetsering) har vi att

$$\xi_1 + \xi_2 \in \text{N}(1.5 + 2.5, \sqrt{0.40^2 + 0.41^2}) = \text{N}(1.5 + 2.5, 0.5728).$$

   Sannolikheten att den totala tiden ligger mellan 3.5 och 4.5 h är

$$\Phi\left(\frac{4.5 - 4.0}{0.5728}\right) - \Phi\left(\frac{3.5 - 4.0}{0.5728}\right) = 2[\Phi\left(\frac{4.5 - 4.0}{0.5728}\right) - 1 = 0.62.$$

   (c) Antag istället att de börjar jobba samtidigt, och att förutsättningarna i övrigt är som ovan. Sannolikheten att Börje blir klar först:

$$\xi_1 - \xi_2 \in \text{N}(1.0, 0.5728),\ P(\xi_1 \geq \xi_2) = 1 - P(\xi_1 - \xi_2 \leq 0) = \Phi\left(\frac{1.0}{0.5728}\right) = 1 - 0.04 = 0.96.$$

4. (2+3 poäng) Talen $(25.0, 26.0, 24.5, 22.0, 25)$ är ett observerat stickprov på en normalfördelning med väntevärde μ och standardavvikelse σ. Räknehjälp: Från stickprovet beräknas medelvärde $\overline{x} = 24.5$ och standardavvikelse $s = 1.5$.

   (a) Ett tvåsidigt 95%:s konfidensintervall för väntevärdet μ. Stickprovets storlek är $n = 5$, så vi behöver kvantilen $t_{4,0.025} = 2.78$. Intervallet skrivet som en olikhet och intervall är
   $$24.5 \pm \frac{t_{5,0.025} \cdot s}{\sqrt{5}} = 24.5 \pm 1.86 \text{ resp. } [22.6, 26.4]$$

   (b) Ett uppåt begränsat 95%:s konfidensintervall för standardavvikelsen σ: Vi behöver kvantilen $\chi^2_{0.95}(4) = 0.71$ så att intervallet för σ^2 och σ är
   $$\left[0, \frac{4s^2}{0.71}\right] = [0, 12.7] \text{ resp. } [0, 3.56].$$

5. (2+4 poäng) Antalet cyklister som passerar GB-bron i Stuvesbergs Centrum under en minut är Poissonfördelat med väntevärde 1.5. Antalet cyklister under olika minuter antas oberoende av varandra.

   (a) Sannolikheten att minst två cyklister passerar bron under en given minut är
   $$P(\xi \geq 2) = 1 - P(\xi = 0) - P(\xi = 1) = 1 - e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!}\right) = 0.44$$

   (b) Låt $\zeta = \sum_{k=1}^{60} \xi_k$. Sannolikheten (approximativt) att minst 60 cyklister passerar bron under en timma är
   $$P(\zeta > 60)$$
   där ζ approximativt är $\mathrm{N}(60\lambda, \sqrt{60\lambda})$ enl. CGS och $60\lambda = 90$, $\sqrt{60\lambda} = 9.48683$.
   $$P(\zeta > 60) = 1 - P(\zeta \leq 60) \approx 1 - \Phi\left(\frac{60 - 90}{9.48683}\right) = \Phi\left(\frac{30}{9.48683}\right) = 0.999$$

6. (3 + 2 + 2 poäng)
   Familjen Yrsel åkte på gemensam cykelsemester förra sommaren. Semestern varade i 32 dagar och varje dag innebar en cykelsträcka på 3 mil (vi kan tänka oss att varje sträcka var likvärdig i svårighetsgrad).
   Vi kan tänka oss att tiden en sträcka cyklades är en responsvariabel (y). Responsvariabeln kan påverkas av tre olika faktorer (ålder, kön och om det regnar eller inte). Då familjen består av jämnåriga föräldrar och deras två barn som är tvillingsyskon (en flicka och en pojke) samt att det regnade 16 av dagarna som familjen cyklade så kan man undersöka responsvariabeln med hjälp av ett 2^3 försök där det existerar 4 replikat för varje provgrupp.

   (a) Beräkna huvud- och samspels-effekter.

Lösning

$$
\begin{aligned}
M &= \frac{164.9 + 43 + 159.1 + 38.5 + 224.7 + 179 + 222.2 + 183}{8} = 151.8 \\
l_A &= \frac{-164.9 + 43 - 159.1 + 38.5 - 224.7 + 179 - 222.2 + 183}{4} = -81.85 \\
l_B &= \cdots = -2.2 \\
l_C &= \cdots = 100.9 \\
l_{AB} &= \cdots = 2.0 \\
l_{AC} &= \cdots = 38.4 \\
l_{BC} &= \cdots = 2.9 \\
l_{ABC} &= \cdots = 1.3
\end{aligned}
$$

   (b) Beräkna ett 95%-igt referensintervall och avgör vilka skattade effekter som är signifikanta. Till din hjälp får du att den skattade standardavvikelsen för effekterna har räknats ut till $s_{\text{effekt}} = 4.5$. Denna har man räknat ut med hjälp av alla s_i värdena i tabellen.

Faktorer	A: Ålder	B: Kön	C: Regn
Låg (-)	12	Man	Torrt
Hög (+)	42	Kvinna	Blött

Nr.	A	B	C	$\hat{y}_i$	s_i
1	-	-	-	164.9	4.3
2	+	-	-	43.0	3.9
3	-	+	-	159.1	4.8
4	+	+	-	38.5	5.1
5	-	-	+	224.7	7.0
6	+	-	+	179.0	4.9
7	-	+	+	222.2	7.7
8	+	+	+	183.0	10.5

Tabell 1: Faktorer samt försöksplan för tiden det tar att cykla en sträcka om 3 mil.

Nr.	A	B	C	AB	AC	BC	ABC	$\hat{y}_i$	s_i
1	-	-	-	+	+	+	-	164.9	4.3
2	+	-	-	-	-	+	+	43.0	3.9
3	-	+	-	-	+	-	+	159.1	4.8
4	+	+	-	+	-	-	-	38.5	5.1
5	-	-	+	+	-	-	+	224.7	7.0
6	+	-	+	-	+	-	-	179.0	4.9
7	-	+	+	-	-	+	-	222.2	7.7
8	+	+	+	+	+	+	+	183.0	10.5

Tabell 2: Provtagningsplan med samspelseffekterna tillagda.

Lösning

Först beräknar vi den gemensamma skattningen av variansen såsom

$$s_p^2 = \frac{3s_1^2 + 3s_2^2 + \dots}{3 * 8} = 40.65.$$

Eftersom effekten har variansen $\sigma^2_{\text{effekt}} = 4\frac{\sigma^2}{8}$ så blir skattningen av effekternas varians

$$s^2_{\text{effekt}} = \frac{40.65}{2} = 20.3.$$

Alltså får effekten den skattade standardavvikelsen $s_{\text{effekt}} = 4.5$. Det tvåsidiga 95%-iga referensintervallet fås genom att multiplicera standardavvikelsen för effekten med 97.5% kvantilen för t-fördelningen med frihetsgraden $8(4-1) = 24$. Alltså blir referensintervallet

$$0 \pm t_{0.025}(df = 24) \cdot s_{\text{effekt}} = 0 \pm 2.06 \cdot 4.5 = 0 \pm 9.27.$$

Från referensintervallet kan vi slutleda oss till att effekterna A, C, och AC är signifikanta eftersom dessa alla har ett absolutvärde större än 9.27.

(c) Under familjen Yrsels cykelsemester så cyklade far och dotter på racingcyklar och mor och son på mountainbikes. Detta kan ha spelat in i hur snabbt de cyklade. Man kan alltså räkna in en fjärde faktor, D som motsvarar cykeltyp där låg (-) skulle motsvara racingcykel och (+) mountainbike.

Försöksplanen är alltså egentligen en reducerad försöksplan, 2^{4-1}-plan. Bestäm generator, ord och upplösning för denna reducerade försöksplan.

Lösning

Då en person är antingen gammal och kvinna eller ung och man så kommer de cykla på mountainbikes, annars inte. Det betyder alltså att D:s kolumn motsvarar tvåfaktorsamspelet AB då AB har positivt värde då A och B är båda låga eller A och B är båda höga.

Nr.	A	B	C	AB	AC	BC	ABC	D	$\hat{y}_i$	s_i
1	-	-	-	+	+	+	-	+	164.9	4.3
2	+	-	-	-	-	+	+	-	43.0	3.9
3	-	+	-	-	+	-	+	-	159.1	4.8
4	+	+	-	+	-	-	-	+	38.5	5.1
5	-	-	+	+	-	-	+	+	224.7	7.0
6	+	-	+	-	+	-	-	-	179.0	4.9
7	-	+	+	-	-	+	-	-	222.2	7.7
8	+	+	+	+	+	+	+	+	183.0	10.5

Tabell 3: Tabell med den exta faktorn D inkluderad.

Man har alltså generatorn $D = AB$. Detta ger den definierande relationen $I = ABD$. Ordet är alltså ABD som har upplösning III. Vi har alltså en 2^{4-1}_{III}-plan.

7. (4 + 2 poäng) Företaget *Måleri AB* tillverkar penslar. I deras fabrik utövar de styrande kontroll genom att kontrollera längden på 10 av de tillverkade penslarna varje dag. Under två arbetsveckor uppmätte de följande stickprovsmedelvärden och stickprovsstandardavvikelser på sina penslar.

Nr.	1	2	3	4	5	6	7	8	9	10
$\bar{x}_i$	12.28	10.41	10.22	11.67	13.41	10.88	8.65	9.49	6.95	8.83
s_i	4.24	3.79	4.37	3.72	3.62	5.46	5.84	4.20	3.52	5.06

Tabell 4: Tabell över prickade värden under 2 arbetsveckor på *Måleri AB*.

(a) Skapa $\bar{x}$- och s-diagram för att avgöra om penseltillverkningen verkar vara under statistisk kontroll under dessa två arbetsveckor.

Lösning

Vi vill skapa ett $\bar{x}-$diagram för att avgöra om medelvärdet förändras under tidsperioden.

$$C_l = \mu \approx \bar{\bar{x}} = \frac{\sum_{i=1}^{10} \bar{x}_i}{10} = 10.28$$

$$S_ö = C_l + 3\frac{\sigma}{\sqrt{10}} \approx c_l + A_3\bar{s} = 10.28 + 0.975 \cdot 4.38 = 10.28 + 4.27 = 14.55$$

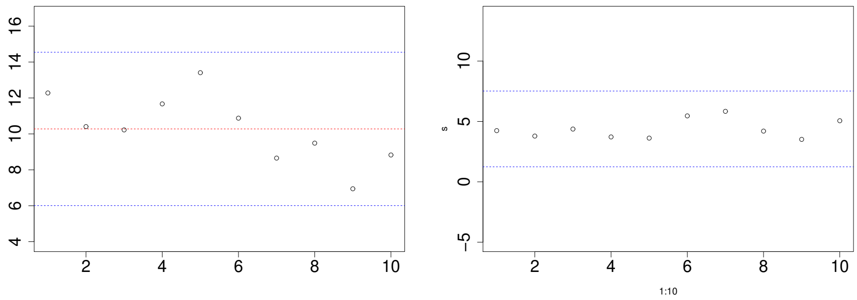
$$S_u = C_l - 3\frac{\sigma}{\sqrt{10}} \approx c_l - A_3\bar{s} = 10.28 - 4.27 = 6.01$$

Vi vill även se om spridningen förändras. Då vi har s_i-värden så får vi göra ett $s-$diagram för spridningen.

$$S_ö \approx B_4\bar{s} = 1.716 \cdot 4.38 = 7.52$$

$$S_u \approx B_3\bar{s} = 0.284 \cdot 4.38 = 1.24$$

Ingen punkt ligger utanför styrgränserna så penslarna ser ut att tillverkas under statistisk kontroll.



(a) $\bar{x}-$diagram (b) $s-$diagram

Figur 1: Styrdiagram för uppgiften.

(b) För att penslarna skall leva upp till *Måleri AB*s kvalitetskrav så måste penslarna ha en längd mellan 8 och 11 centimeter. Beräkna korrigerat duglighetsindex och avgör om processen producerar penslar som med tillräckligt god marginal lever upp till kvalitetskraven.

Lösning

Duglihetsindex är

$$C_p = \frac{T_{\ddot{o}} - T_u}{6\sigma} \approx \frac{T_{\ddot{o}} - T_u}{6\bar{s}} = \frac{11-8}{6 \cdot 4.38} = 0.114.$$

Då C_p bör vara större än 1.33 så ser vi redan nu att spridningen är för stor för att vi skall vara nöjda med kvaliteten på penslarna. Eftersom det frågas om korrigerat duglihetsindex så fortsätter vi ändå att räkna ut detta,

$$CM = 2\frac{|\mu - M|}{T_{\ddot{o}} - T_u} \approx 2\frac{|\bar{x} - M|}{T_{\ddot{o}} - T_u} = 2\frac{10.28 - 9.5}{3} = \frac{1.56}{3} = 0.52.$$

Det korrigerade duglighetsindexet blir därför

$$C_{pk} \approx C_p(1 - 0.52) = 0.114 \cdot 0.48 = 0.055.$$

Eftersom $0.055 < 1.33$ så lever inte tillverkningsprocessen upp till de krav som våra toleransgränser kräver.

8. (1 + 2 + 2 + 3 poäng)

JosAB är ett företag som köper in äpplejuice och säljer i Göteborgsregionen. De vill utföra acceptanskontroll på de partier de köper in för att minska antalet sålda förpackningar med missfärgningar.

Varje parti de köper in innehåller 2000 förpackningar med juice. JosAB anser att om mindre än 50 av dessa är missfärgade så är partiet så bra att de med stor sannolikhet bör accepteras för försäljning. Om mer än 200 av förpackningarna är missfärgade så anses partiet vara så dåligt att det med stor sannolikhet bör avvisas.

(a) Räkna ut acceptabel kvalitetsnivå (d.v.s p_1) och gränskvalitet (d.v.s p_2).

Lösning

$p_1 = \frac{50}{2000} = 2.5\%$ och $p_2 = \frac{200}{2000} = 10\%$.

(b) JosAB vill ha en producentrisk (α) på 5% och en konsumentrisk (β) på 10%. Vilken enkel provtagningsplan bör de använda för att minimera antalet kontrollerade förpackningar men samtidigt ha tillräckligt små konsument- och producentrisker? (Använd bifogat binomialfördelningsnomogram.)

Lösning

Genom binomialnomogrammet i figure 2 ser vi att $n \approx 75$ och $c \approx 4$ ger en bra enkel provtagningsplan (egentligen bör n vara 78 eller 79 om $C = 4$ men alla värden mellan 70 och 80 godtas).

(c) Konstruera en dubbel provtagningsplan som uppfyller samma kriterier.

Lösning

Låt oss titta på kvoten $\frac{p_2}{p_1} = \frac{10}{2.5} = 4 \approx 3.88$. Närmast värdet från tabell hittar vi alltså i provtagningsplanen med $n_1 = n_2, c_1 = 2, c_2 = 5$ och $r_1 = r_2 = 6$.

Antal enheter att kontrollera kan man få ut på två sätt från tabellen:

$$n \cdot p_1 = 1.43 \Leftrightarrow n = \frac{1.43}{0.025} = 57.2$$

$$n \cdot p_2 = 5.55 \Leftrightarrow n = \frac{5.55}{0.1} = 55.5.$$

Vi kan ta det största av dess: $n_1 = n_2 = 57.2 \approx 58$. Vi tog det största av dessa och avrundade uppåt eftersom vi vill välja en plan som fortfarande uppfyller att riskerna är tillräckligt små.

Om du ej löst (c) så välj en dubbel provtagningsplan som du tycker är rimlig och räkna på den i fråga (d) istället.

(d) Räkna ut producentrisken med den givna dubbla provtagningsplanen. Ange approximationer om sådana behöver göras.

Lösning

Vi kan approximera antalet funna defekta med en binomialfördelning. Antalet defekta i första urvalet kan då modelleras med en slumpvariabel $\xi \sim Bin(n = 58, p)$ och för andra urvalet med $\xi_2 \sim Bin(n = 58, p)$.

Producentrisken kan räknas ut som

$$\alpha = 1 - L(p_1) = 1 - \mathbb{P}(\xi_1 \leq 2; p = p_1) - \sum_{i=3}^{5} \mathbb{P}(\xi_2 \leq 5 - \xi_1 | \xi_1 = i; p = p_1)\mathbb{P}(\xi_1 = i; p = p_1).$$

För att ta fram dessa sannolikheter behöver vi först räkna ut sannolikheten att ξ_1 antar värdena $0, 1, 2, 3, 4, 5$.

$$\mathbb{P}(\xi_1 = 0) = \binom{58}{0} p_1^0 (1 - p_1)^{58} = 0.23$$

$$\mathbb{P}(\xi_1 = 1) = \binom{58}{1} p_1^1 (1 - p_1)^{58-1} = 0.34$$

$$\mathbb{P}(\xi_1 = 2) = \binom{58}{2} p_1^2 (1 - p_1)^{58-2} = 0.25$$

$$\mathbb{P}(\xi_1 = 3) = \binom{58}{3} p_1^3 (1 - p_1)^{58-3} = 0.12$$

$$\mathbb{P}(\xi_1 = 4) = \binom{58}{4} p_1^4 (1 - p_1)^{58-4} = 0.04$$

$$\mathbb{P}(\xi_1 = 5) = \binom{58}{5} p_1^5 (1 - p_1)^{58-5} = 0.01$$

Sannolikheten att acceptera i första urvalet är

$$\mathbb{P}(\xi_1 = 0) + \mathbb{P}(\xi_1 = 1) + \mathbb{P}(\xi_1 = 2) = 0.23 + 0.34 + 0.25 = 82\%.$$

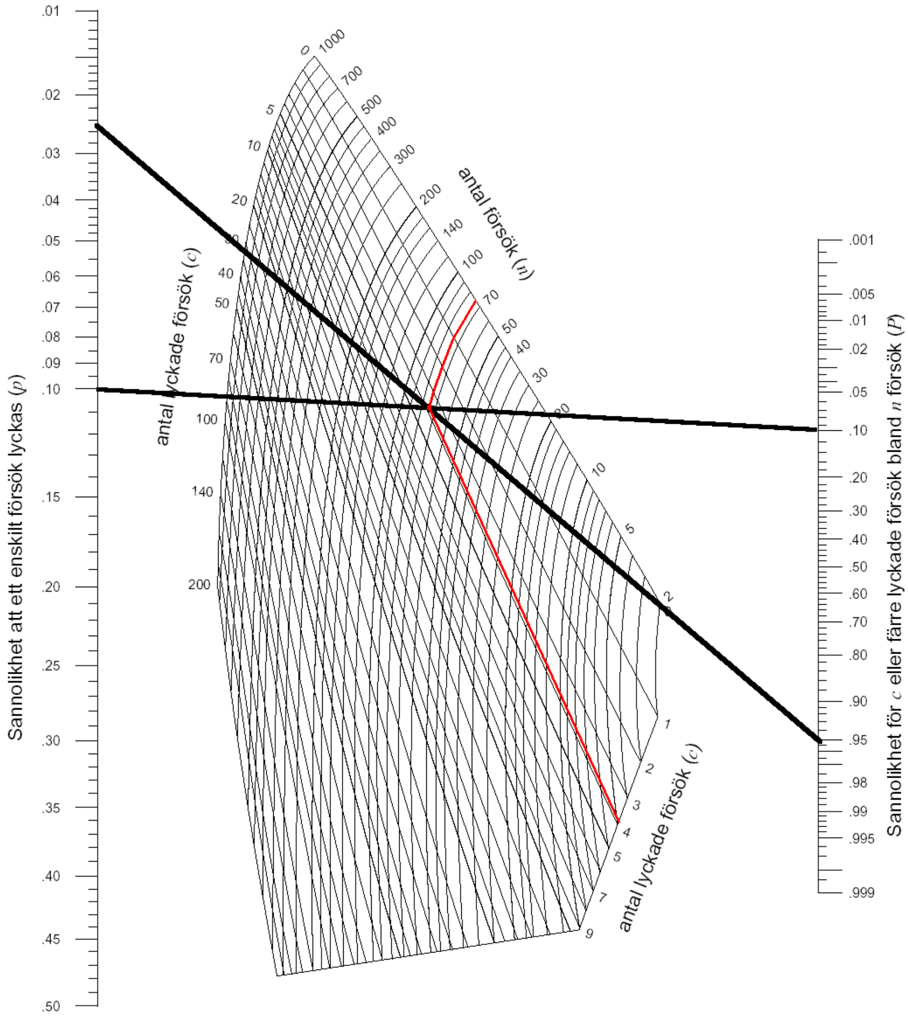
Sannolikheten att acceptera i andra urvalet är

$$\begin{aligned} &(\mathbb{P}(\xi_2 = 0) + \mathbb{P}(\xi_2 = 1) + \mathbb{P}(\xi_2 = 2))\,\mathbb{P}(\xi_1 = 3) + \\ &(\mathbb{P}(\xi_2 = 0) + \mathbb{P}(\xi_2 = 1))\,\mathbb{P}(\xi_1 = 4) + \mathbb{P}(\xi_2 = 0)\mathbb{P}(\xi_1 = 5) \\ &= 0.82 \cdot 0.12 + 0.57 \cdot 0.04 + 0.23 \cdot 0.01 = 12\% \end{aligned}$$

Slår man ihop dem får man en sannolikhet på 94% att acceptera ett parti om $p = p_1$. Det motsvarar en producentrisk på 6% (uträknat i datorn utan avrundning av decimalerna blir det 5.147%). Producentrisken är alltså lite högre än vad vi egentligen ville tillåta. Detta beror på att tabellen vi använde är approximativ.

NOMOGRAM ÖVER BINOMIALFÖRDELNINGEN

$P = P(X \le c)$ där $X \in Bin(n, p)$; X = antal lyckade försök



Figur 2: Ifyllt binomialnomogram för lösningen.