Tillämpad matematisk statistik LMA521 (EPI/Design/Maskin/Mekatronik-programmen) Tentamen 2018-06-05

Tid: 14.00-18.00. **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.**
Kursansvarig: Reimond Emanuelsson/Johan Tykesson
Telefonvakt/jour: Reimond, 0708948456 / Johan, 0703182096
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 4.1, 4.2, 4.1 och 4.3 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 $\sigma = 0.1$.

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

   Lösning:

   (a)
   $$\bar{x} \pm 1.96\sigma/\sqrt{n} = 4.175 \pm 1.96 \times 0.1/2 = 4.175 \pm 0.098 = [4.077, 4.273].$$

   (b)
   $$\bar{x} \pm 3.18 s/\sqrt{n} = 4.175 \pm 3.18 \times 0.096/2 = 4.175 \pm 0.153 = [4.022, 4.328].$$

2. (4+4 poäng) Antag att livslängden för en viss typ av smart-phone är exponentialfördelad med väntevärde 2 år. Antag att vi har 100 smartphones av denna modell, och att deras livslängder kan antas oberoende av varandra.

   (a) Beräkna (approximativt) sannolikheten att summan av de 100 telefonernas livslängder är mer än 210 år.

   (b) Beräkna (approximativt) sannolikheten att antalet telefoner som fungerar efter 2 år är mindre än eller lika med 53.

   Lösning:

   (a) Låt ξ_i vara livslängden för smartphone nummer i, $i = 1, \ldots, 100$. Det gäller att $E(\xi) = 2$ och $\sigma(\xi) = 2$. Enligt CGS gäller att $T = \sum_{i=1}^{100}$ är approximativt $N(100 \times 2, \sqrt{100} \times 2) = N(200, 20)$. Alltså blir
   $$P(T > 210) = 1 - P(T \leq 210) = 1 - P(\frac{T - 200}{20} \leq \frac{210 - 200}{20})$$
   $$\approx 1 - \Phi(0.5) = 1 - 0.691 = 0.309.$$

(b)
$$P(\xi \geq 2) = \int_2^\infty \frac{1}{2} e^{-x/2}\, dx = [-e^{-x/2}]_2^\infty = e^{-1} \approx 0.368.$$
Alltså är antalet hela smartphones efter 2 år binomialfördelat med $n = 100$ och $p = 0.368$. Beteckna detta antal med η. Eftersom $np(1-p) \approx 23.25 > 10$ så är η approx $N(np, \sqrt{np(1-p)}) = N(36.8, 4.82)$. Därfor blir
$$P(\eta \leq 53) = P(\frac{\eta - 36.8}{4.88} \leq \frac{53 - 36.8}{4.82}) \approx \Phi(3.36) \approx 0.9996. \quad (1)$$

3. (4 poäng) Antag att A, B och C är händelser, och att de alla är oberoende av varandra. Antag också att det gäller att $P(A) = P(B) = P(C) = 0.4$.
   (a) Vad är sannolikheten att alla händelserna inträffar?
   (b) Vad är sannolikheten att exakt en av händelserna inträffar?

Lösning:

(a) Oberoende ger att
$$P(A \cap B \cap C) = 0.4^3 = 0.064.$$
(b) Notera att
$$P(\text{ endast } A) = P(A \cap B^c \cap C^c) = P(A)P(B^c)P(C^c) = 0.4 \times 0.6^2 = 0.144,$$
där oberoende användes i andra likheten. På samma sätt fås att $P(\text{ endast } B) = 0.144$ och $P(\text{ endast } C) = 0.144$. Så
$$P(\text{exakt en av händelserna inträffar}) = 3 \times 0.144 = 0.432.$$

4. (2+2+2 poäng) Antag att ξ är en kontinuerlig stokastisk variabel med frekvensfunktion
$$f(x) = \begin{cases} 5x^4 & \text{för } 0 \leq x \leq 1 \\ 0 & \text{för övrigt} \end{cases}$$
   (a) Visa att $f(x)$ är en frekvensfunktion.
   (b) Beräkna väntevärde och standardavvikelse för ξ.
   (c) Beräkna den betingade sannolikheten
$$P(0.2 \leq \xi \leq 0.6 \,|\, 0.4 \leq \xi \leq 0.8).$$

Lösning:

(a) För $0 \leq x \leq 1$ gäller att $5x^4 \geq 0$. Dessutom är $\int_0^1 5x^4\, dx = 1$.

(b)
$$E(\xi) = \int_0^1 5x^5\, dx = \frac{5}{6}$$
.
$$E(\xi^2) = \int_0^1 5x^6\, dx = \frac{5}{7}.$$
$$\sigma = \sqrt{5/7 - (5/6)^2} \approx 0.141.$$

(c)

$$P(0.2 \le \xi \le 0.6 \,|\, 0.4 \le \xi \le 0.8) = \frac{P(0.4 \le \xi \le 0.6)}{P(0.4 \le \xi \le 0.8)} = \frac{\int_{0.4}^{0.6} 5x^4\,dx}{\int_{0.4}^{0.8} 5x^4\,dx} \approx 0.213.$$

5. (2+1+2+1 poäng) Ett fullständigt faktorförsök har gjorts enligt försöksplanen i Tabell 1. Två olika faktorer användes och i varje försöksgrupp gjordes 20 mätningar. I tabellen kan man se stickprovsmedelvärdet och stickprovsvariansen från varje försöksgrupp.

Försöksgrupp nr	A	B	AB	Resultat $\hat{y}$	Resultat s^2
1	-	-	+	$\bar{y}_1 = 66.81$	$s_1^2 = 333.61$
2	+	-	-	$\bar{y}_2 = 48.87$	$s_2^2 = 303.71$
3	-	+	-	$\bar{y}_3 = 59.78$	$s_3^2 = 262.26$
4	+	+	+	$\bar{y}_4 = 57.04$	$s_4^2 = 441.22$

Tabell 1: Provtagningsplan och uträknade stickprovsmedelvärden och stickprovsvarianser.

(a) Beräkna huvudeffekterna, medelvärdet samt tvåfaktorsamspelet.

(b) Man har beräknat att ett 95% referensintervall ges av $[-8.15, 8.15]$. Vilka huvudeffekter anser du är signifikanta (med en signifikansgrad av 5%) givet detta referensintervall?

(c) Om vi antar att vi skulle vara intresserade av ytterligare en faktor, C, och vi väljer provtagningsplanen såsom i Tabell 2 nedan, vad blir alias för A, B och C?

(d) Vad är upplösningen för den reducerade försöksplanen?

Försöksgrupp nr	A	B	C
1	-	-	-
2	+	-	-
3	-	+	+
4	+	+	+

Tabell 2: Reducerad provtagningsplan.

Lösning:

(a)

$$\text{medelvärdet} = \frac{\bar{y}_2 + \bar{y}_4 + \bar{y}_1 + \bar{y}_3}{4} = \frac{48.87 + 57.04 + 66.81 + 59.78}{4} = 58.13 \tag{2}$$

$$l_A = \frac{\bar{y}_2 + \bar{y}_4}{2} - \frac{\bar{y}_1 + \bar{y}_3}{2} = \frac{48.87 + 57.04 - 66.81 - 59.78}{2} = -10.34 \tag{3}$$

$$l_B = \frac{\bar{y}_3 + \bar{y}_4}{2} - \frac{\bar{y}_1 + \bar{y}_2}{2} = \frac{59.78 + 57.04 - 66.81 - 48.87}{2} = 0.57 \tag{4}$$

$$l_{AB} = \frac{\bar{y}_1 + \bar{y}_4}{2} - \frac{\bar{y}_2 + \bar{y}_3}{2} = \frac{66.81 + 57.04 - 48.87 - 59.78}{2} = 7.6 \tag{5}$$

$$\tag{6}$$

(b) Huvudeffekten av A anser vi är signifikant eftersom den är utanför referensintervallet. Huvudeffekt B och tvåfaktorsamspelet AB är för nära 0.

(c) Kolumnerna för B och C är identiska. Alltså har vi generatorn $B = C$. Detta ger oss den definierande relationen $I = BC$ vilket i sin tur ger oss sammanblandningsmönstret:

$$A = ABC \tag{7}$$

$$B = C \tag{8}$$

$$AB = AC \tag{9}$$

$$BC = I \tag{10}$$

(d) Upplösningen är längden på det kortaste ordet, alltså 2.

6. (2+2+3 poäng) I en fabrik produceras skåpsluckor (av samma sort) på två band, A och B. På band A produceras 25% fler luckor än på band B (det vill säga, på samma tid som det produceras 100 luckor på band B, produceras 125 luckor på band A.) Andelen defekta luckor från band A är 2% och från band B 3.5%.

(a) Vad är sannolikheten att en skåpslucka är defekt?

(b) Vad är betingade sannolikheten att en skåpslucka kommer från band A, givet att den är defekt?

(c) Var är betingade sannolikheten att en skåpslucka kommer från band B, givet att den är korrekt (ej defekt)?

Lösning: Låt D beteckna händelsen att luckan defekt.

(a)

$$\begin{cases} P(A) + P(B) = 1 \\ P(A) = 1.25P(B) \end{cases} \iff \begin{cases} P(A) = 5/9 \\ P(B) = 4/9 \end{cases}$$

$$P(D) = P(D|A)P(A) + P(D|B)P(B) = 002 \cdot 5/9 + 0.035 \cdot 4/9 = 2/75 \approx 0.027.$$

(b)

$$P(A|D) = P(D|A) \cdot \frac{P(A)}{P(D)} = 0.02 \cdot \frac{5/9}{2/75} \approx 0.42.$$

Sannolikheten, givet att den är defekt, att den kommer från band A är 0.42.

(c)

$$P(B|D^c) = \frac{(1 - P(D|B))P(B)}{1 - P(D)} = \frac{0.965}{73/75} \cdot \frac{5}{9} \approx 0.44.$$

7. (1.5+1.5+1+2+1 poäng) Parasollföretaget SolOchBad AB utför styrande kontroll för att kontrollera om tygarean av tillverkade parasoller börjar avvika från deras från början välkalibrerade produktionsprocess. De är intresserade av att använda sig av ett medelvärdesdiagram och ett R-diagram. Därför väljer de slumpmässigt ut 5 parasoller varje vecka och mäter tygarean på var och en av dem. Från dessa fem mätningar så räknar de ut ett stickprovsmeelvärde, $\bar{x}_i$, och en variationsbredd, R_i.

Vecka (i)	1	2	3	4	5	6	7	8	9	10
$\bar{x}_i$	1.48	1.57	1.66	1.38	1.42	1.72	1.48	1.47	1.51	1.48
R_i	0.54	0.27	0.43	0.37	0.21	0.28	0.42	0.18	0.37	0.28
Vecka (i)	11	12	13	14	15	16	17	18	19	20
$\bar{x}_i$	1.45	1.38	1.34	1.44	1.41	1.49	1.56	1.52	1.40	1.61
R_i	0.41	0.30	0.62	0.75	0.39	0.65	0.34	0.73	1.06	0.84

Tabell 3: Tabell över stickprovsmedelvärden och variationsbredder för alla 20 veckor.

(a) Räkna ut styrgränser (kontrollgränser) för ett medelvärdesdiagram ($\bar{x}$-diagram).

(b) Räkna ut styrgränser (kontrollgränser) för ett R-diagram.

(c) Anser du att processen är under statistisk kontroll? Motivera din åsikt med hjälp av styrgränserna i dina diagram.

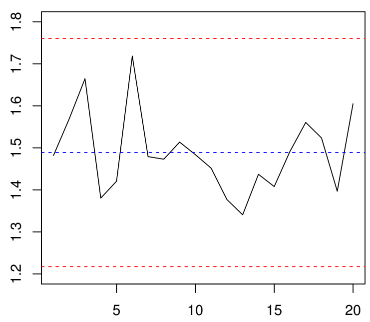
(d) SolOchBad AB har övre och undre toleransgränser för tygarean som är 1.5 och 1.7. Om vi antar att processen skulle varit under statistisk kontroll och att σ skattas med $\frac{\bar{R}}{d_2}$ där $d_2 = 2.326$, vad är det (skattade) korrigerade kapablitetsindexet (duglighetsindexet)?

(e) Vad bör man dra för slutsats av värdet på det korrigerade kapabilitetsindexet?

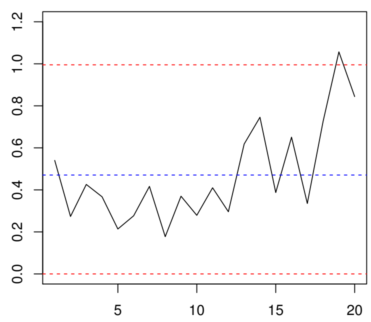
Lösning:

(a) Eftersom det var 5 mätningar per vecka så tittar vi på rad 5 i tabellen för konstruktion av kontrolldiagram. Då vi skall skapa et medelvärdesdiagram givet datan vi mätt och variationsbredden så vill vi använda oss av konstanten $A_2 = 0.577$.
Vi räknar ut centrallinjen $\bar{\bar{y}} = \frac{1.48+1.57+...+1.61}{20} = 1.49$. Vi räknar även ut medelvärdet av variationsbredderna, $\bar{R} = \frac{0.54+0.27+...0.84}{20} = 0.47$. Den övre styrgränsen blir nu

$$S_{\ddot{o}} = \bar{\bar{y}} + A_2\bar{R} = 1.49 + 0.577 \cdot 0.47 = 1.76.$$



Figur 1: $\bar{x}$-diagram



Figur 2: R-diagram.

Figur 3: Styrdiagram för parasollernas tygarea.

Likaså blir den under gränsen

$$S_{\mathrm{u}} = 1.49 - 0.577 \cdot 0.47 = 1.22.$$

Detta motsvarar medelvärdesdiagrammet såsom i Figur 3.

(b) För att räkna ut stygränserna till R-diagrammet använder vi oss av D_3 och D_4, fortfarande på rad 5 i tabellen.

$$S_{\mathrm{u}} = D_3\bar{R} = 2.115 \cdot 0.47 = 0 \quad (11)$$
$$S_{\mathrm{ö}} = D_4\bar{R} = 2.115 \cdot 0.47 = 0.99 \quad (12)$$

R-diagrammet kan ses i Figur 3.

(c) Processen är inte under statistisk kontroll. Vecka 19 så är variationsbredden ovanför den övre styrgränsen i R-diagrammet.

(d) Vi har alltså fått att $M = 1.6$, $T_{\mathrm{ö}} = 1.7$ och $T_{\mathrm{u}} = 1.5$. Vidare kan vi räkna ut att $\sigma = \frac{\bar{R}}{d_2} = \frac{0.47}{2.326} = 0.202$. Kapabilitetsindex är då

$$C_p = \frac{T_{\mathrm{ö}} - T_{\mathrm{u}}}{6\sigma} = \frac{0.2}{6 \cdot 0.202} = 0.165.$$

$CM = 2\frac{|M-\mu|}{T_{\mathrm{ö}}-T_{\mathrm{u}}} \approx 2\frac{|M-\bar{\bar{y}}|}{0.2} = 10(1.6 - 1.49) = 1.1$. Därför blir det korrigerade kapabilitetsindexet

$$C_{pk} = C_p(1 - CM) = -0.0165.$$

(e) Eftersom $C_p < 1.33$ så är spridningen i processen för stor. Eftersom C_{pk} är väsentligt mindre än C_p så verkar det som att processen dessutom är dåligt centrerad.

8. (2+2+2 poäng) En pluton bestående av 30 soldater delas helt slumpmässigt in i 10 grupper om vardera 3 personer. I varje grupp tilldelas en person rollen som befäl, en person rollen som spanare, och en person rollen som skytt. Detta sker också helt slumpmässigt. Tre av soldaterna heter Kurt, Sven och Veronika.

   (a) Vad är sannolikheten att Kurt, Sven och Veronika hamnar i samma grupp? Svara på exakt form.
   (b) Vad är sannolikheten att Kurt blir spanare, Sven blir skytt och Veronika blir befäl? (De behöver inte hamna i samma grupp)
   (c) Vad är sannolikheten att Kurt och Sven hamnar i samma grupp, men Veronika hamnar i en annan grupp?

Lösning:

(a)

$$
\begin{aligned}
P(\text{K,S,V i samma grupp}) &= 10 \times P(\text{K,S,V i första gruppen}) \\
&= 10 \times \frac{3}{30} \times \frac{2}{29} \times \frac{1}{28} = \frac{1}{406}
\end{aligned}
$$

(b)

$$
P(\text{K spanare, S skytt, V befäl}) \quad = \quad \frac{10}{30} \times \frac{10}{29} \times \frac{10}{28} \quad = \quad \frac{25}{609}.
$$

(c)

$$
P(\text{K,S i samma grupp, V i annan grupp})
$$

$$
= 10 \times P(\text{K,S i första gruppen, V i annan grupp}) = 10 \times \frac{3}{30} \times \frac{2}{29} \times \frac{27}{28} = \frac{27}{406}
$$

Lycka till!