Avd. Matematisk statistik

KTH Matematik

TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING FÖR K OCH B ONSNDAGEN DEN 21 APRIL 2004 KL 8.00–13.00.

Examinator: Gunnar Englund, 790 7416.

Tillåtna hjälpmedel: Formel- och tabellsamling i Statistik med försöksplanering. Kalkylator.

Varje korrekt löst uppgift ger 10 poäng så maximalpoängen blir 50 poäng. För godkänt krävs 20 poäng. Resultatet anslås senast onsdagen den 12 maj 2004 på matematisk statistiks anslagstavla i entréplanet, Lindstedtsvägen 25.

Införda beteckningar skall förklaras och definieras. Resonemang och räkningar skall vara så utförliga att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet.

Lycka till!

- -

Uppgift 1

a) Den stokastiska variabeln X har väntevärde 3 och variansen 2. Bestäm, med lämplig approximation $D(Xe^X)$. (5 p)

b) Låt X och Y vara oberoende normalfördelade stokastiska variabler. X är $\mathrm{N}(3, 2)$ och Y är N(2,5 , 1,5). Beräkna $P(X > 2 \text{ och } Y > 3)$ samt $P(X + Y > 5)$. (5 p)

Uppgift 2

Volymen av blod som strömmar från hjärtat efter en hjärtmuskelsamman- dragning kallas slagvolymen. För en individ som vilar är slagvolymen i medeltal cirka 75 ml. Vid en medicinsk undersökning av sambandet mellan slagvolym och ålder fick man följande data.

Ålder	20	25	30	35	40	45	50	55	60	65	70
Slagvolym	74	76	77	74	71	72	70	68	67	64	62

a) Ansätt en lämplig modell för data. (2 p)

b) Testa om åldern har någon signifikant inverkan på slagvolymen. 5% signifikansnivå. (5 p)

c) Beräkna ett 95% konfidensintervall för förväntad slagvolym hos en 50-åring. (3 p)

Uppgift 3

Tre termometrar används regelbundet i ett laboratorium. För att kontrollera den relativa noggrannheten hos termometrarna placerades dessa i slumpmässig ordning i en cell som höll 0 ^{o}C. Varje termometer placerades i cellen fyra gånger, med följande resultat.(^{o}C).

Termometer	1	2	3
	0.10	-0.20	0.90
	0.90	0.80	0.20
	-0.80	-0.30	0.30
	-0.20	0.60	-0.30

Hjälpsumma:$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = 3.3267$ a) Ange lämplig statistisk modell samt testa på 5%-nivån om termometrarna mäter likvärdigt. (6 p)

b) Reducera eventuellt modellen samt beräkna ett 95% konfidensintervall för den första termometerns systematiska fel. (4 p)

Uppgift 4

En industrikemist studerar hur utbytet av en kemisk process beror av halten av två i processen deltagande ämnen A och B. Vart och ett av de två ämnena prövar han i fyra olika koncentrationer och för varje kombination av A- och B-nivåer gör han en bestämning. I tabellen nedan presenteras för varje faktorkombination bestämningarna.

	B1	B2	B3	B4	$\bar{y}_{i.}$
A1	0.433	0.467	0.501	0.505	0.4765
A2	0.413	0.448	0.475	0.496	0.4580
A3	0.422	0.445	0.473	0.487	0.45675
A4	0.400	0.438	0.487	0.499	0.4560
$\bar{y}_{.j}$	0.4170	0.4495	0.4840	0.49675	$\bar{y}_{..} = 0.4618125$

Hjälpsumma: $\sum_i \sum_j (y_{ij} - \bar{y}..)^2 = 0.0172064$.

a) Ange en statistisk modell samt testa om ämnet A har en signifikant effekt. Gör detsamma avseende B. Använd 5% signifikansnivå. (6 p)

b) Beräkna ett 95% konfidensintercvall för den systematiska skillnaden mellan halten B4 och B1 av ämne B. (4 p)

Uppgift 5

I tabellen nedan redovisas resultatet från ett försök där man varierade fyra faktorer A,B,C och D vardera på två nivåer (kodade som + och –).

A	B	C	D	Mätdata
−	−	−	−	71
+	−	−	−	61
−	+	−	−	90
+	+	−	−	82
−	−	+	−	68
+	−	+	−	61
−	+	+	−	87
+	+	+	−	80
−	−	−	+	61
+	−	−	+	50
−	+	−	+	89
+	+	−	+	83
−	−	+	+	59
+	−	+	+	51
−	+	+	+	85
+	+	+	+	78

Ur dessa data räknade man fram följande skattningar:

$$\widehat{I} = 72.25,\ \widehat{A} = -4,\ \widehat{B} = 12,\ \widehat{C} = -1.125,\ \widehat{D} = -2.75,$$

$$\widehat{AB} = 0.5,\ \widehat{AC} = 0.375,\ \widehat{AD} = 0,\ \widehat{BC} = -0.625,\ \widehat{BD} = 2.25,\ \widehat{CD} = -0.125,$$

$$\widehat{ABC} = -0.375,\ \widehat{ABD} = 0.25,\ \widehat{ACD} = -0.125,\ \widehat{BCD} = -0.375,\ \widehat{ABCD} = -0.125.$$

a) Skaffa Dig en skattning av försöksfelsvariansen genom att anta att 3- och 4-faktorsamspelen är försumbara. Beräkna också medelfelet för huvudeffekterna och 2-faktorsamspelen. (5 p)

b) Dra slutsatser om vilka huvudfaktorer och 2-faktorsamspel som är signifikanta (nivå 5%). (5 p)

LÖSNINGAR TILL TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING FÖR K ONSDAGEN DEN 21 APRIL 2004 KL 8.00-13.00

Uppgift 1

a) Enligt Gauss' approximationsformler får vi $E(g(X)) \approx g(E(X))$ och $V(g(X)) \approx (g'(E(X)))^2 V(X)$ och vi har här $g(x) = xe^x$ och erhåller alltså $g'(x) = (1+x)e^x$. Detta ger $E(Xe^X) \approx E(X)e^{E(X)} = 3e^3 \approx \underline{60.26}$ och $V(Xe^X) \approx ((1+3)e^3)^2 \cdot 2$ som ger $D(Xe^X) \approx \sqrt{16e^6} \cdot \sqrt{2} \approx \underline{113.62}$.

b)

$$P(X > 2 \text{ och } Y > 3) = (\text{p g a oberoendet}) = P(X > 2)P(Y > 3) =$$

$$P\left(\frac{X-3}{2} > \frac{2-3}{2}\right) P\left(\frac{Y-2.5}{1.5} > \frac{3-2.5}{1.5}\right) =$$

$$(1 - \Phi(-0.5))\,(1 - \Phi(1/3)) \approx \Phi(0.5)\,(1 - \Phi(0.33)) \approx 0.6915 \cdot (1 - 0.6293) \approx \underline{0.256}.$$

Vidare är $X+Y$ normalfördelad eftersom X och Y är det och de är oberoende. $E(X+Y) = E(X) + E(Y) = 3 + 2.5 = 5.5$ och $V(X+Y) = V(X) + V(Y) = 2^2 + 1.5^2 = 6.25$ dvs $D(X+Y) = \sqrt{6.25} = 2.5$ och vi ser att $X+Y$ är $N(5.5,\, 2.5)$ vilket ger

$$P(X+Y > 5) = P\left(\frac{X+Y-5.5}{2.5} > \frac{5-5.5}{2.5}\right) = 1 - \Phi(-0.2) = \Phi(0.2) \approx 0.5793 \approx \underline{0.580}.$$

Uppgift 2

a) Ansätt en linjär regressionsmodell med ålder som oberoende variabel x och slagvolymen som den beroende variabeln y.

b) Beräkningar ger $\bar{x} = 45 \quad \bar{y} = 70.4545 \quad \sum_i (x_i - \bar{x})^2 = 2750 \quad \sum_i (y_i - \bar{y})^2 = 232.7273$ och $\sum_i (x_i - \bar{x})y_i = -755$. Härav fås att, med sedvanliga beteckningar, att $\beta^* = \sum_i (x_i - \bar{x})y_i / \sum_i (x_i - \bar{x})^2 = -0.2745$, $\alpha^* = \bar{y} - \beta^* \bar{x} = 82.81$ och $s^2 = \frac{1}{11-2}\big(\sum_i (y_i - \bar{y}) - \beta^{*2} \sum_i (x_i - \bar{x})^2 = 2.8273$
Ett 95% konfidensintervall för β ges av

$$\beta^* \pm t_{0.025}(9) s / \sqrt{\sum_i (x_i - \bar{x})^2} = \underline{-0.2745 \pm 0.0725}$$

Eftersom 0 inte tillhör intervallet förkastas hypotesen att åldern ej skulle ha någon inverkan på slagvolymen, dvs slagvolymen har en signifikant påverkan på slagvolymen.

c) Ett 95 % konfidensintervall ges av

$$\alpha^* + \beta^* 50 \pm t_{0.025}(9)s\sqrt{\frac{1}{11} + \frac{(50-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}} = \underline{69.08 \pm 1.20}$$

Uppgift 3

a) Ensidig variansanalys. Låt y_{ij}=observation nr j från termometer nr i. Då fås $\bar{y}_{1.} = 0$ $\bar{y}_{2.} = 0.225$ $\bar{y}_{3.} = 0.3143$ och $\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = 0.04042$. Härav fås variansanalystabellen

Källa	Fg	Kvs	Mkvs
Mellan stermometrar	2	0,1717	0,0858
Inom termometrar	9	3,1550	0,3506
Totalt	11	3,3267	

Testkvot $F_{termometer} = \dfrac{0.0858}{0.3506} = 0.2448 < F_{0.05}(2,9) = 2.26$ och således kan man inte förkasta hypotesen att termomterarna mäter likvärdigt.

b) Vi reducerar modellen och antar att termometrarna är likvärdiga. Nu σ^2-skattning blir 3.3267/11 och konfidensintervall för systematiska felet $\bar{y}_{..} \pm t_{0.025}(11)s/\sqrt{12} = \underline{0.17 \pm 1,16}$.

Uppgift 4

a) Tvåsidig variansanalys. Vi antar att faktorerna är additiva. Beräkningar ger variansanalystabellen

Källa	Fg	Kvs	Mkvs
Mellan A-nivåer	3	0,001159	0,000386
Mellan B-nivåer	3	0,01549	0,00516
Residual	9	0,000557	0,0000619=$\widehat{\sigma}^2$
Totalt	15	0,017206	

Testkvoter $F_A = 0.000386/0.0000619 = 6.2 > F_{0.05}(3,9) = 3.86$ och $F_B = 0.00516/0.0000619 = 82 > F_{0.05}(3,9)$ vilket visar att både A- och B-faktorerna är signifikanta.

b) Konfidensintervall ges av

$$\bar{y}_{.4} - \bar{y}_{.1} \pm t_{0.025}(9)\widehat{\sigma}\sqrt{1/4 + 1/4} = \underline{0.07975 \pm 0.013}$$

Uppgift 5

Om 3- och 4-faktorsamspelen sätts till 0 erhåller vi

$$\frac{\widehat{\sigma^2}}{16} = \frac{1}{5}\left(\widehat{ABC}^2 + \widehat{ABD}^2 + \widehat{ACD}^2 + \widehat{BCD}^2 + \widehat{ABCD}^2\right) = 0.075$$

och vi får medelfelen $\dfrac{\widehat{\sigma}}{\sqrt{16}} \approx 0.274$ och detta gäller för samtliga skattningar. Vidare är $\widehat{\sigma^2} = 1.2$

b) Vår σ-skattning är baserad på 5 frihetsgrader och vi får därför konfidensintervall av typen skattning$\pm t_{0.025}(5)0.274 \approx 2.57 \cdot 0.274 \approx 0.704$ och detta ger att A,B,C,D och BD-effekterna är signifikanta på 5%-nivån.