



KUNGL
TEKNISKA
HÖGSKOLAN

Matematiska institutionen
avd matematisk statistik

TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING
TISDAGEN DEN 22 APRIL 2003 KL 14.00 – 19.00.

Examinator: Gunnar Englund, tel. 7907416

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

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng.

Resultatet anslås senast tisdagen den 13 maj 2003 på Matematisk statistiks anslags-tavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten.

Lycka till!

Uppgift 1

245 kast görs med en välgjord tärning (dvs lika sannolikhet för de sex möjliga utfallen). Beräkna sannolikheten att relativa frekvensen sexor avviker med mer än 0.05 (åt något håll) från idealvärdet $1/6$. Väl motiverade approximationer är tillåtna.
(10 p)

Uppgift 2

Vid en undersökning av försurningen i naturen ville man studera om försurningen var olika kraftig i olika miljöer. Man valde därför ut 2 sjöar som huvudsakligen låg vid åkermark respektive skogsmark. Ett antal prover togs från varje sjö och analyserades med samma metod, varvid följande pH-värden erhöles:

Sjö 1: x_1, x_2, x_3, x_4, x_5 där $\sum_1^5 x_k = 27$ och $\sum_1^5 x_k^2 = 146.12$

Sjö 2: y_1, y_2, y_3, y_4 där $\sum_1^4 y_k = 20$ och $\sum_1^4 y_k^2 = 100.20$

Analysresultaten antas vara utfall av oberoende normalfördelade stokastiska variabler med väntevärden m_1 (hörande till sjö 1) resp m_2 (hörande till sjö 2) och gemensam standardavvikelse σ .

a) Beräkna ett exakt 95% konfidensintervall för $m_1 - m_2$.
(7 p)

- b) Testa på nivån 5% hypotesen $H_0 : m_1 = m_2$ mot alternativet $H_1 : m_1 \neq m_2$. Det skall klart framgå huruvida H_0 förkastas eller ej. (3 p)

Uppgift 3

För att jämföra fyra olika att torka trä, torkades att antal träbitar medelst metoderna och fuktigheten efter torkning mättes. Resultat. (i %)

Metod 1	23.6	22.5	21.3	24.7	24
Metod 2	26	24.8	24.2	23.5	26.2
Metod 3	27.3	26.9	26.1	27.7	25.7
Metod 4	26.6	25.5	20.3	22.8	24.8

Totalkkvs: $\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = 74.5175$ Totalmedelvärde: $\bar{y}_{..} = 24.725$

- a) Ställ upp lämplig statistisk modell och testa hypotesen att det inte är någon skillnad mellan torkningsmetoderna. 5 % signifikansnivå. (5 p)
- b) Ge 95 % konfidensintervall för de parvisa förväntade skillnaderna i torkningsresultat. (5 p)

Uppgift 4

För att jämföra effekten av fyra olika halter av en viss ingrediens i en kemisk process, kördes processen under var och en av veckans fem arbetsdagar med var och en av halterna. Under varje arbetsdag randomiserade man ordningen av de olika haltkörningarna. Utbytet mättes och man erhöll följande resultat:

	Halt 1	Halt 2	Halt 3	Halt 4
Dag 1	53	54	55	54
Dag 2	52	53	57	55
Dag 3	51	54	55	57
Dag 4	53	55	56	53
Dag 5	54	56	55	57

Totalkkvs: $\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = 52.95$ Totalmedelvärde: $\bar{y}_{..} = 54.45$

Ställ upp en lämplig statistisk modell. Undersök om det är någon signifikant variation i utbytet mellan olika dagar. Testa sedan (efter eventuell modellreduktion) hypotesen att utbytet är oberoende av val av aktuell halt. 5 % signifikansnivå. (10 p)

Uppgift 5

- a) A, B, C och D är fyra faktorer i ett 2^{4-1} -försök. D är en ren blockfaktor, som man på goda grunder antar inte samspelar med någon annan faktor. Man vill använda detta faktum för att lägga upp försöket med en koppling som möjliggör att alla andra faktorer kan skattas. Tre- och fyrfaktorsamspelet antas som vanligt försumbara. Ange en sådan koppling, och ange hur alla faktorer är kopplade med varandra. (3 p)

- b) Följande observationer erhöles, i standardordning som om observationerna kom från ett 2^3 -försök med faktorer A, B och C, dvs i ordningen

---, +- -, -+ -, +++-, --+, +- +, -+++, ++++

85.5 75.1 93.2 145.4 83.7 77.6 95 141.8

Räknehjälp: Om M är matrisen

$$\begin{pmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

och y datavektorn ovan är (797,3 82,5 153,5 115,5 -1,1 -1,1 -2,5 -9,7)

$$y \cdot M = (797.3 \quad 82.5 \quad 153.5 \quad 115.5 \quad -1.1 \quad -1.1 \quad -2.5 \quad -9.7)$$

Skatta effekterna.

(3 p)

c) Man har även gjort fyra observationer i en centrpunkt och fått resultaten

101.3 102.8 102.5 101.8

Använd dessa data för att skatta σ^2 och sedan testa om effekterna är signifikant skilda från 0.

(4 p)



KUNGL
TEKNISKA
HÖGSKOLAN

Matematiska institutionen
avd matematisk statistik

LÖSNINGAR TILL TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING
TISDAGEN DEN 22 APRIL 2003 KL 14.00 – 19.00.

Uppgift 1

Låt X vara antalet sexor vid de 245 kasten. Då är X

$$Bin(245, 1/6) \approx N\left(\frac{245}{6}, \sqrt{245 \cdot \frac{1}{6} \cdot \frac{5}{6}}\right) \text{ - förd. eftersom } 245 \cdot \frac{1}{6} \cdot \frac{5}{6} = 34.0 > 10.$$

Detta ger

$$P\left(\left|\frac{X}{245} - \frac{1}{6}\right| > 0.05\right) = P\left(\frac{|X - \frac{245}{6}|}{\sqrt{245 \cdot \frac{1}{6} \cdot \frac{5}{6}}} > 0.05 \frac{\sqrt{245}}{\sqrt{\frac{1}{6} \cdot \frac{5}{6}}}\right) \approx$$
$$2\left(1 - \Phi\left(0.05 \frac{\sqrt{245}}{\sqrt{\frac{1}{6} \cdot \frac{5}{6}}}\right)\right) = 2(1 - \Phi(2.1)) = \underline{0.036}$$

Uppgift 2

Vi har situationen "Två oberoende stickprov". Vi får på sedvanligt sätt konfidensintervall

$$\bar{x} - \bar{y} \pm t_{0.025}(5+4-2)s\sqrt{\frac{1}{5} + \frac{1}{4}} = \bar{x} - \bar{y} \pm 2.36s\sqrt{\frac{1}{5} + \frac{1}{4}}.$$

Vi får $\bar{x} = 27/5 = 5.4$, $\bar{y} = 20/4 = 5.0$ och vidare

$$s_x^2 = \frac{1}{5-1} \left(146.12 - \frac{27^2}{5}\right) = 0.32/4, \quad s_y^2 = \frac{1}{4-1} \left(100.20 - \frac{20^2}{4}\right) = 0.20/3$$

som ger

$$s^2 = \frac{4s_x^2 + 3s_y^2}{4+3} = (0.32 + 0.20)/7 = 0.52/7 = 0.0743$$

och vi får konfidensintervallet

$$5.4 - 5 \pm 2.36\sqrt{0.0743}\sqrt{\frac{9}{20}} \approx \underline{0.40 \pm 0.43}.$$

b) Eftersom 0 ingår i konfidensintervallet kan vi ej förkasta H_0 , dvs ingen skillnad kan anses vara statistiskt säkerställd.

Uppgift 3

a) Vi får följande variansanalystabell. Totalkvadratsumma och totalmedelvärde är givna. Stickprovsmedelvärden är

$\bar{y}_1 = 23.22$, $\bar{y}_2 = 24.94$, $\bar{y}_3 = 26.74$ och $\bar{y}_4 = 24$. Det ger oss följande variansanalystabell.

Källa	Fg	Kvs	Mkvs
Mellan metoder	3	34.485	11.49517
Inom metoder	16	40.032	$2.50 = \hat{\sigma}^2$
Totalt	19	74.517	

F-kvot är $F_{metod} = \frac{11.5}{2.502} = 4.594391 > 3.24$. Signifikant skillnad mellan torkningsmetoder.

b) Konfidensintervall för $\mu_i - \mu_j$ ges av $\bar{y}_i - \bar{y}_j \pm t_{0.025}(16)\hat{\sigma}\sqrt{\frac{1}{5} + \frac{1}{5}} = \bar{y}_i - \bar{y}_j \pm 2.473 \cdot 1.00 = \bar{y}_i - \bar{y}_j \pm 2.47$. Vi får intervallen

$$1 - 2: -1.72 \pm 2.47$$

$$1 - 3: -3.52 \pm 2.47$$

$$1 - 4: -0.78 \pm 2.47$$

$$2 - 3: -1.80 \pm 2.47$$

$$2 - 4: 0.94 \pm 2.47$$

$$3 - 4: 2.74 \pm 2.47.$$

Uppgift 4

Variansanalysmodell, tvåsidig indelning, 1 obs/cell, additiv modell antas. Faktorer dag och halt. Dag slumpmässig faktor, halt systematisk.

Vi får följande variansanalystabell. Totalkvadratsumma och totalmedelvärde är givna. Rad- och kolumnmedelvärden är

$\bar{y}_1 = 54$, $\bar{y}_2 = 54.25$, $\bar{y}_3 = 54.25$, $\bar{y}_4 = 54.25$, $\bar{y}_5 = 55.5$, $\bar{y}_1 = 52.6$, $\bar{y}_2 = 54.4$, $\bar{y}_3 = 55.6$, $\bar{y}_4 = 55.2$.

Således är $4 \cdot \sum_i (y_i - \bar{y}_i)^2 = 5.7$ och $5 \cdot \sum_j (y_j - \bar{y}_j)^2 = 26.55$. Härav erhålles variansanalystabellen

Källa	Fg	Kvs	Mkvs
Mellan dag	4	5.7	1.425
Mellan halt	3	26.55	8.85
Residual	12	20.7	1.73
Totalt	19	52.95	

F-kvot $F_{dag} = \frac{1.425}{1.73} = 0.83 < F_{0.05}(4,12) = 3.26$. Ej signifikant variation mellan dagar och vi väljer att reducera modellen. Lagg ihop kvadratsummorna dag och residual och dess frihetsgrader och vi erhåller ny variansskattning $\hat{\sigma}^2 = \frac{5.7 + 20.7}{4 + 12} = 1.65$

F-kvot är $F_{halt} = \frac{8.850}{1.65} = 5.36 > F_{0.05}(3,16) = 3.24$. Signifikant skillnad mellan halter.

Uppgift 5

a) Använd kopplingen D+ABC. Vi har då kopplingarna I+ABCD, A+BCD, B+ACD, C+ABD, AB+CD, AC+BD, AD+BC. Eftersom trefaktorsamspelet är försumbara och D inte samspekar kan alla andra effekter skattas.

b) Skattningarna erhålles genom att dividera $y \cdot M$ med 8. Vi får skattningarna (obs ABC=D, ABC försumbar effekt)

Effekt	I	A	B	AB	C	AC	BC	D(=ABC)
Skattning	99.6625	10.3125	19.1875	14.4375	-0.1375	-0.1375	-0.3125	-1.2125

c) Variansen σ^2 skattas med stickprovsvariansen $s^2=0.46$. Den har 3 frihetsgrader varför ett konfidensintervall för en effekt ges av skattning $\pm t_{0.025}(3)s/\sqrt{8}$ =skattning ± 0.76 . A, B, och AB och D är signifikanta.