

TENTAMEN: Statistisk modellering för I3, TMS160, tisdagen den 29 mars 2005 kl 8.30-11.30 på V. **Jour:** Holger Rootzén , ankn. 3578, mobil 0730 794222

Hjälpmedel: Utdelad formelsamling med tabeller, BETA, på kursen använd ordlista och typgodkänd räknedosa.

Poängberäkning: Uppgifterna är av flervalstyp, där endast ett alternativ är rätt. Korrekt besvarad uppgift ger 2 poäng, obesvarad uppgift (vet inte eller alternativ f) ger 0 poäng och felaktigt besvarad uppgift ger -0.5 poäng (flera ifyllda alternativ ger automatiskt -1/2 poäng). Inlämnade lösningar kommer ej tas hänsyn till vid rättningen. Fyll i och lämna in denna sida.

Svar: Lägg ut i studieportalen efter tentamens slut.

Uppgift	a	b	c	d	e	f (vet ej)	Poäng
1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
13	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
14	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

- 1 Nedanstående tabell visar antalet besökande (uppmätt som antalet övernattningar) för tio skidsportsorter under en viss period. För varje ort anges dessutom den samlade pistlängden (i miles) samt samt liftkapaciteten (i personer per timme).

Ort	Pist- längd	Lift - kapa- citet	Antal övernatt- ningar
1	10.5	2 200	19 929
2	2.5	1 000	5 839
3	13.1	3 250	23 696
4	4.0	1 475	9 881
5	14.7	3 800	30 011
6	3.6	1 200	7 241
7	7.1	1 900	11 634
8	22.5	5 575	45 684
9	17.0	4 200	36 476
10	6.4	1 850	12 068

Man vill använda linjär regressionsanalys till att beskriva hur antalet övernattningar hänger samman med pistlängd och liftkapacitet.

En eventuell multikollinearitet mellan pistlängd och liftkapacitet ger problem därför att

- a Variansen av intercepttermen blir mycket stor.
- b Man kan inte skilja effekten av pistlängden och liftkapaciteten åt.
- c Residualkvadratsumman blir för stor.
- d Man får ett litet värde på R^2
- e Inget av ovanstående.
- f Vet ej.

2 Man har samlat in 20 observationer från en fördelning med okänd varians. Vilken av följande faktorer påverkar *inte* bredden på ett konfidensintervall för väntevärdet.

- a Addition av 37 till den första observationern.
- b Storleken på stickprovet.
- c Storleken av skattningen av standardavvikelsen.
- d Konfidensgraden.
- e Alla ovanstående alternativ påverkar längden av konfidensintervallet.
- f Vet ej.

- 3 Ett företag som tillverkar en bestämd produkt startar tillverkningen när den mottar en order. Storleken av ordern varierar från gång till gång. Man är intresserad av att beskriva hur produktionstiden beror av orderstorleken, och vill använda linjär regressionsanalys till detta.

Tabellen nedan visar orderstorlek (i antal tillverkade enheter) och produktionstid (i mantimmar) för ett antal order

Order nr	Orderstorlek	Produktionstid
1	30	73
2	20	50
3	60	128
4	80	170
5	40	87
6	50	108
7	60	135
8	30	69
9	70	148
10	60	132

Man kan kontrollera modellens normalfördelningsantagande genom att rita en Q-Q plott av

- a Antalet mantimmar.
- b Antalet mantimmars avvikelser från regressionslinjen.
- c De standardiserade värdena av antalet mantimmar
- d Orderstorleken.
- e Inget av ovanstående.
- f Vet ej.

4 En studie av 1436 kvinnor som varit gifta minst en gång gav följande resultat.

Utbildning	Gifta en gång	Gifta flera gånger	Totalt
Högskola	550	61	611
Ingen Högskola	681	144	825
Totalt	1231	205	1436

Vad kan man som starkast säga om hypotesen “antalet gånger kvinnorna varit gifta är oberoende av deras utbildning”?

- a Vi kan förkasta på signifikansnivå 10%
- b Vi kan förkasta på signifikansnivå 5%
- c Vi kan förkasta på signifikansnivå 2.5%
- d Vi kan förkasta på signifikansnivå 1%
- e Inget av ovanstående.
- f Vet ej.

5 Ett test har utvecklats för att hitta en sjukdom bland individer som är äldre än 50 år. Vi vet att approximativt 10% av den åldersgruppen är smittade. Vi vet även att vid en undersökning av smittade personer gav testet ett korrekt resultat i 85% av fallen samt att vid en undersökning av friska människor rapporterade testet att 4% var smittade.

Vad är den avrundade sannolikheten att en individ har sjukdomen givet att testet indikerar detta?

- a 0.76
- b 0.78
- c 0.80
- d 0.12
- e Inget av ovanstående.
- f Vet ej.

6 För att mäta något man är intresserad av är det vanligt att man först genomför en referensmätning. Vi antar att man gör oberoende mätningar. Först gör man 25 mätningar på referensämnet och beräknar standardavvikelsen till 1.08 g/ml.

Därefter gör man 12 mätningar av halten i det prov man verkligen är intresserad av. Medelvärdet av dessa mätningar blev 91.12 och standardavvikelsen blev 0.73 g/ml

Man vet att standardavvikelsen är den samma vid referensmätningar och riktiga mätningar.

Det mest korrekta 95%- iga konfidensintervallet för halten i provet ges då av:

a $91.12 \pm t_{35,0.975} \cdot \sqrt{\frac{1}{12} \cdot \frac{24 \cdot 1.08^2 + 11 \cdot 0.73^2}{35}}$ [g/ml]

b $91.12 \pm t_{35,0.95} \cdot \sqrt{\frac{1}{12} \cdot \frac{24 \cdot 1.08^2 + 11 \cdot 0.73^2}{43}}$ [g/ml]

c $91.12 \pm t_{11,0.975} \cdot \sqrt{\frac{0.73^2}{15}}$ [g/ml]

d $91.12 \pm t_{11,0.95} \cdot \sqrt{\frac{0.73^2}{15}}$ [g/ml]

e Inget av ovanstående

f vet ej

7 Vad undersöks med randomized complete block design (RBCD)?

- a Medelvärdet av k stycken populationer som också påverkas av en extern störvariabel.
- b Om en mängd observationer är dragna av en speciell fördelning.
- c Linjärt beroende mellan variabler.
- d Effekten av k stycken behandlingsformer på en homogen mängd.
- e Inget av ovanstående.
- f Vet ej.

8 Vi kategoriserar frukostflingor efter sockerhalt. Vi låter kategori 0 stå för flingor av müslityp, 1 för majsflingor och 2 för flingor av typen honungsrostadechokladdoppadesockerpuffar. 90 butiker valdes ut slumpmässigt. I 30 butiker så undersöktes flingorna närmast kassorna, i 30 butiker flingorna i mitten och i 30 butiker flingorna längst bort från kassorna.

Man vill undersöka om sockerinnehållet har ett samband med placeringarna relativt kassorna. Hur gör man det bäst?

- a Ensidig variansanalys utan blockindelning (one-way model).
- b Ensidig variansanalys med blockindelning (RCBD)
- c Regressionsanalys.
- d χ^2 -test för oberoende
- e Inget av ovanstående.
- f vet inte

9 Den stokastiska variabeln X har väntevärde 2 och standardavvikelse 3 och variabeln Y har väntevärde 3 och standardavvikelse 4. Vidare är korrelationen mellan X och Y lika med 0.25. Sätt $Z = X - Y + 5$.

Då är variansen för Z lika med

- a 11
- b 13
- c 16
- d 18
- e Inget av ovanstående.
- f vet ej

10 Vilka antaganden bygger F-testet i en multipel linjär regressionsanalys på?

- 1: Residualerna är normalfördelade.
- 2: Residualerna är oberoende.
- 3: Värdena av prediktorerna är stationära.

Svar:

- a Antagande 1 och 3, men inte 2.
- b Antagande 2 och 3, men inte 1.
- c Alla tre antaganden (1-3).
- d Antagande 1 och 2, men inte 3.
- e Inget av ovanstående.
- f Vet ej.

11 Tabellen nedan visar ANOVA-tabellen för en tvåsidig variansanalys.

Analysis of variance				
Source	DF	Sum of squares	Mean square	F Stat
A	*	20.48	*	?
B	1	1.70	1.70	*
A×B	*	157.25	78.62	*
Error	18	112.05	6.23	
Total	23	291.48		

Värdena i några av fälten saknas och är markerade med (*). Från de givna siffrorna kan man ändå beräkna värdet av F-statistikan (markerad med ett frågetecken i tabellen). Den blir:

- a 10.24/6.23
- b 10.24/112.05
- c 20.48/6.23
- c 20.48/112.05
- e 20.48/291.48
- f Vet ej.

- 12 I syfte att undersöka en (kostsam) mätprocedur vid tre olika laboratorier delades ett provmaterial in i 9 delar och 3 delar skickades till vardera av de tre laboratorierna. I respektive laboratorium bad man vidare att de tre provmaterialen skulle mätas av olika laboratorieassistenter för att också få med denna variation i undersökningen.

Ett test om mätproceduren ger samma resultat vid de tre laboratorierna görs bäst med:

- a Regressionsanalys.
- b Ensidig variansanalys med blockindelning (RCBD).
- c Ensidig variansanalys utan blockindelning (one-way model).
- d χ^2 -test för oberoende i en kontingenstabell.
- e Inget av ovanstående.
- f Vet ej.

- 13 Ett jordprovs absorption av fosfat kan användas som ett mått på effektiviteten av pesticider och andra kemikalier.

Tabellen nedan visar sammanhängande mätvärden av fosfatabsorptionsindex och mängden järn och aluminium i ett antal jordprov.

Järn	Aluminium	Absorptionsindex
175	21	18
111	24	14
124	23	18
130	64	26
173	38	26
169	33	21
169	61	30
160	39	28
244	71	36
257	112	65
333	88	62
199	54	40

Man gjorde en linjär regressionanalys för att beskriva hur fosfatabsorptionen beror av järn- och aluminiumhalt i jorden. Valda delar av output ges nedan.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2806.216517	1403.108258	65.84	<.0001
Error	9	191.783483	21.309276		
Corrected Total	11	2998.000000			

Parameter Estimates					
Variable	DF	Estimate	Std Error	t Stat	Pr > t
Intercept	1	-7.31	4.38	-1.67	0.1298
järn	1	0.11	0.03	3.36	0.0084
alum	1	0.34	0.07	4.64	0.0012

Det predikterade värdet av av absorptionsindex vid ett järninnehåll på 170 och ett aluminiuminnehåll på 25 blir

a $-7.31 + 0.11 \times 170 + 0.34 \times 25$

b $-7.31 + 0.34 \times 170 + 0.11 \times 25$

c $4.38 + 0.03 \times 170 + 0.07 \times 25$

d $173 + 0.11 \times 170 + 0.34 \times 25$

e $-7.31 - 4.38 + (0.11 - 0.03) \times 170 + (0.34 - 0.075) \times 25$

f Vet ej.

14 I en undersökning var man intresserad av hur skördeutfallet för en nyförädlad växt beror av ett antal uppmätta faktorer. Som ett led i analysen gjorde man ett antal interaktionsgrafer.

1 Hur kan man upptäcka interaktion i en interaktionsgraf?

2 I vilka av följande fall kan man förvänta sig interaktion: a) mellan regnmängd och antalet soltimmar, och b) ögonfärg och hårfärg hos traktorföraren?

Dessa frågor besvaras bäst med:

a 1: Parallella linjer. 2: b) men inte a).

b 1: Korsande linjer. 2: a) och b).

c 1: Parallella linjer. 2: ingen av a) och b).

d 1: Korsande linjer. 2: a) men inte b).

e Inget av ovanstående.

f Vet ej.

- 15 En firma säljer swimmingpooler, spas och bastuar. Ägaren beslutar sig för att undersöka om åldern av försäljningspersonalen (grupperad i “20-29-åriga”, “30-39-åriga”, “40-49”-åriga och “50-åriga och över”) och om produkttypen (“swimmingpool”, “spa” och “sauna”) har något inflytande på den månatliga försäljningen.

Resultatet visas i tabellen nedan. Man vet att där är två observationer i varje cell.

Variation	SS
ålder	168.033
produkttyp	1762.067
växelverkan	7955.267
fel	2574.000
total	12459.367

F-testkvantiteten för test av hypotesen att det inte finns någon växelverkan mellan ålder och produkttyp är

- a $(168.033/3)/(2574.000/7)$
- b $(168.033/3)/((7955.267 + 2574.000)/13)$
- c $(168.033/3)/(7955.267/6)$
- d $(7955.267/6)/(2574.000/12)$
- e Inget av ovanstående.
- f Vet ej.