TENTAMEN: Statistisk modellering för I3, TMS160, fredagen den 26 Augusti kl ? på ?. **Jour:** Holger Rootzén, ankn. 3578
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: Läggs ut i studieportalen efter tentamens slut.

Uppgift	a	b	c	d	e	f (vet ej)	Poäng
1	☐	☐	☐	☐	☐	☐	
2	☐	☐	☐	☐	☐	☐	
3	☐	☐	☐	☐	☐	☐	
4	☐	☐	☐	☐	☐	☐	
5	☐	☐	☐	☐	☐	☐	
6	☐	☐	☐	☐	☐	☐	
7	☐	☐	☐	☐	☐	☐	
8	☐	☐	☐	☐	☐	☐	
9	☐	☐	☐	☐	☐	☐	
10	☐	☐	☐	☐	☐	☐	
11	☐	☐	☐	☐	☐	☐	
12	☐	☐	☐	☐	☐	☐	
13	☐	☐	☐	☐	☐	☐	
14	☐	☐	☐	☐	☐	☐	
15	☐	☐	☐	☐	☐	☐	

1 Tre olika leukemibehandlingar jämfördes i ett försök där 63 patienter slumpmässigt indelades i tre lika stora grupper som behandlades med var sin metod. Efter behandlingsperiodens slut bedömde man om behandlingen haft någon effekt, eller om man inte kunde påvisa någon effekt.

Behandling	Effekt	Ingen effekt
1	12	9
2	17	4
3	16	5

Försöket analyseras bäst med

a Regressionsanalys

b Ensidig variansanalys utan blockindelning (one-way model)

c Ensidig variansanalys med blockindelning (RCBD)

d Ett Pearson χ^2-test

e Inget af ovanstående

f Vet inte

2 Som ett led i ett studium av relationen mellan kroppsvikt och hjärtvikt hos däggdjur fångade man 19 murmeldjur og bestämte deres kroppsvikt och hjärtvikt i gram.

Tabellen nedan ger några grundläggande statistikor för kroppsvikten body och hjärtvikten heart.

body

Moments

N	19.0000	Sum Wgts	19.0000
Mean	3103.6842	Sum	58970.0000
Std Dev	1030.7196	Variance	1062382.89
Skewness	1.2068	Kurtosis	0.9321
USS	202147150	CSS	19122892.1
CV	33.2096	Std Mean	236.4633

heart

Moments

N	19.0000	Sum Wgts	19.0000
Mean	11.9105	Sum	226.3000
Std Dev	1.5776	Variance	2.4888
Skewness	0.9943	Kurtosis	1.0560
USS	2740.1500	CSS	44.7979
CV	13.2453	Std Mean	0.3619

Vidare har man beräknat

$$\sum_{i=1}^{19}(x_i - \overline{x})(y_i - \overline{y}) = -2120.2368,$$

där x_i och $\overline{x}$ är kroppsvikten för det i'te murmeldjuret och medelvärdet av kroppsvikten hos de 19 djuren, och motsvarande är y_i och $\overline{y}$ hjärtvikten och den genomsnittliga hjärtvikten.

Pearson-korrelationen mellan kroppsvikt och hjärtvikt för desssa mätningar ska beräknas som:

a $\dfrac{3103.6842 \cdot 11.9105}{18 \cdot 1030.7196 \cdot 1.5776}$

b $\dfrac{3103.6842 \cdot 11.9105}{18 \cdot 236.4633 \cdot 0.3619}$

c $\dfrac{-2120.2368}{18 \cdot 1030.7196 \cdot 1.5776}$

d $\dfrac{-2120.2368}{1030.7196 \cdot 1.5776}$

e $\dfrac{-2120.2368}{18 \cdot 3103.6842 \cdot 11.9105}$

f Vet inte

3 Varför bör $np_i \geq 5$, $i = 1, ..., c$, i en en-vägs tabell, och hur många frihetsgrader har $X^2 = \sum_{i=1}^{c} \frac{\left(Y_i - np_i^{(0)}\right)^2}{np_i^{(0)}}$ under $H_0 : p_i = p_i^{(0)}$?

(a) ☐ För att undvika beroenden då antalet frihetsgrader ä r c.

(b) ☐ För att centrala gränsvärdessatsen ska ge att $\frac{\left(Y_i - np_i^{(0)}\right)^2}{np_i^{(0)}}$ är χ^2-fördelad med en frihetsgrad för varje i, och $c - 1$.

(c) ☐ För att centrala gränsvärdessatsen ska ge att $\frac{Y_i - np_i^{(0)}}{\sqrt{np_i^{(0)}}}$ är χ^2-fördelad med en frihetsgrad för varje i, och $c - 1$.

(d) ☐ För att centrala gränsvärdessatsen ska ge att $\frac{\left(Y_i - np_i^{(0)}\right)^2}{np_i^{(0)}}$ är χ^2-fördelad med en frihetsgrad för varje i, och c.

(e) ☐ Inget av ovanstående.

(f) ☐ Vet ej.

4 Vad har tolerans och förklaringsvärde gemensamt, och vad indikerar låg tolerans?

(a) ☐Toleransen för regressorn i är 1− förklaringsvärdet för vad man får om man tar den regressorn som respons och gör regression med avseende på övriga regressorer. Låg tolerans indikerar multikolinjäritet.

(b) ☐ Ingenting. Låg tolerans indikerar liten correlation mellan regressorerna.

(c) ☐ Toleransen för regressorn i är 1− förklaringsvärdet för vad man får om man tar den regressorn som respons och gör regression med avseende på övriga regressorer. Låg tolerans indikerar liten correlation mellan regressorerna.

(d) ☐ Ingenting. Låg tolerans indikerar multikolinjäritet.

(e) ☐ Inget av ovanstående.

(f) ☐ Vet ej.

5 Betrakta figurerna 1-4. I vilka figurer är det rimligt att anta att beroendet mellan datamängderna kan beskrivas (rimligt) väl genom att beräkna korrelationskoefficienten?

(a) ☐ I figurerna 1,2 och 3.

(b) ☐ I figurerna 1,2 och 4.

(c) ☐ Endast i figurerna 1 och 2.

(d) ☐ Endast i figur 4.

(e) ☐ Inget av ovanstående.

(f) ☐ Vet ej.

ftbpF6.4931in5.232in0innfplot1.epsftbpF6.606in5.232in0innfplot2.epsftbpF6.4931in5.232in0

6 En pappersmaskin lägger på en tunn skyddsfilm av plast på pappret för att göra det motståndskraftigt mot mekaniskt slitage. Man har undersökt tre olika produktionsomgångar och lagt på filmen med fyra olika koncentrationer(2%, 3%, 4% og 5%) av ett tillsatsmedel. I ett kontrollerat experiment fick man följande styrkemätningar:

	Koncentration			
Produktion	2%	3%	4%	5%
A	10.4	10.2	9.4	9.3
B	11.2	10.5	10.4	10.0
C	11.3	11.0	10.4	9.7

Data analyserades med ensidig variansanalys med blockindelning (RCBD). Tukey's test for additivitet gav följande resultat:

```
TUKEYS TEST FOR ADDITIVITY:
F STATISTIC:    0.16638        DEGREES OF FREEDOM:     1    5
P-VALUE:         0.70023
```

Man kan av detta dra slutsatsen att

a Man bør använda en multiplikativ model

b Analysen ger en klar indikation på att skillnaden mellan effekterna av koncentrationerna av tillsatsmedel beror av vilken produktionsomgång man betraktar.

c Analysen ger ingen indikation på att skillnaden mellan effekterna av koncentrationerna av tillsatsmedel beror av vilken produktionsomgång man betratktar.

d Man bør göra en logaritmisk transformation av de uppmätta styrkorna.

e Man bør göra en logaritmisk transformation af de använda koncentrationerna

f Vet inte

7 Under en studie från 1987 av Gastwirth framkom det att sannolikheten att en lögndetektor tror att man ljuger om man verkligen ljuger är 0.88 samt att sannolikheten att lögndetektorn tror att man talar sanning givet att man verkligen gör det är 0.86. Om man dessutom lyckades mäta att sannolikheten att man talar sanning givet att även lögndetektorn tycker så är 0.94 kan man bakvägen räkna ut sannolikheten att en person ljög. Vad blir då denna sannolikhet, alltså sannolikheten en person ljög?

a ☐ 0.01

b ☐ 0.13

c ☐ 0.03

d ☐ 0.05

e ☐ Inget av ovanstående.

f ☐ Vet ej.

8 Varför kan det vara önskvärt att använda multipel jämförelse istället för ett F-test?

a ☐ Vi får ut mer information i en multipel jämföresle.

b ☐ Det är ej önskvärt.

c ☐ En multipel jämförelse är enklare att genomföra.

d ☐ Multipel jämförelse mäter varians bättre.

e ☐ Inget av ovanstående.

f ☐ Vet ej.

9 Vid en analys prövade man både linjär och icke-linjär regression och fick nästan likadana R^2-värden i båda analyserna. Vilken typ av modell borde du då rekomendera, och hur borde du motivera ditt val?

a ☐ Den icke-linjära för den beräkningsmässigt mindre krävande.

b ☐ Den linjära för den förklarar variationen på ett bättre sätt.

c ☐ Den icke-linjära för den förklarar variationen på ett bättre sätt.

d ☐ Den linjära för den är mindre komplicerad.

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 stationära.

3: Värdena av svarsvariabeln är stationära.

Svar:

a ☐ Antagande 1 och 2, men inte 3.

b ☐ Antagande 1 och 3, men inte 2.

c ☐ Antagande 2 och 3, men inte 1.

c ☐ Alla tre antaganden (1-3).

e ☐ Inget av de ovanstående

f ☐ Vet inte.

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

d ☐ 20.48/6.23

e ☐ 20.48/112.05

e ☐ 20.48/291.48

f ☐ vet inte.

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.

Resultatet visas i följande diagram.

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

a ☐ regressionsanalys.

b ☐ ensidig variansanalys utan blockindelning (one-way model).

c ☐ ensidig variansanalys med blockindelning (RCBD).

d ☐ χ^2-test för oberoende i en kontigenstabell.

e ☐ inget av de ovanstående.

f ☐ vet inte.

13 Ett jordprovs adsorption av fosfat kan användas som ett mått på effektiviteten av pesticider og andra kemikalier.

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

Järn	Aluminium	Adsorptions-index
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 fosfatadsorptionen 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 adsorptionsindex vid ett järninnehåll på 170 och ett aluminiuminnehåll på 25 blir

a -7.31 + 0.11 × 170 + 0.34 × 25

b -7.31 + 0.34 × 170 + 0.11 × 25

c 4.38 + 0.03 × 170 + 0.07 × 25

d 173 + 0.11 × 170 + 0.34 × 25

e -7.31 - 4.38 + (0.11 - 0.03) × 170 + (0.34 - 0.075) × 25

f Vet inte

14 I ett lantbruksförsök var man intresserad av skördeutbytet under olika förhållanden. Betrakta följande frågor:

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

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

c ☐ 1: Korsande linjer. 2: a) men inte b)

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

e ☐ Inget av ovanstående

f ☐ vet inte

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 $(7955.267/6)/(2574.000/12)$

b $(168.033/3)/(2574.000/7)$

c $(168.033/3)/((7955.267 + 2574.000)/13)$

d $(168.033/3)/(7955.267/6)$

e $(1762.067/2)/(2574.000/7)$

f Vet inte