

**ÖVNINGSTENTAMEN:** Statistisk modellering för I3, TMS160

**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 ej) 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"/>	<input type="checkbox"/>	
2	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	
4	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	
6	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	
8	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	
10	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	
12	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	
14	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	

1 I ett laboratorium vill man bestämma ett konfidensintervall för en mätmetod. Man har mycket stor erfarenhet av metoden och vet att den är mycket stabil, men att spridningen beror på en komponent som ofta byts ut.

För att bestämma spridningen måste man därför göra ett antal mätningar på ett referensämne varje gång komponenten bytts ut. Man vet att spridningen är oberoende av väntevärdet av mätningarna och alltså bara beror på komponenten. Först gör man 25 mätningar på referensämnet och beräknar standardavvikelsen till 1.07 g/ml.

Därefter gör man 12 mätningar av halten i det prov man verkligen är intresserad av. Medelvärdet av dessa mätningar blev 101.12 och standardavvikelsen blev 1.13 g/ml

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

a   $101.12 \pm t_{11,0.95} \cdot \sqrt{\frac{1.13^2}{12}}$  [g/ml]

b   $101.12 \pm t_{11,0.975} \cdot \sqrt{\frac{1.13^2}{12}}$  [g/ml]

c   $101.12 \pm t_{35,0.95} \cdot \sqrt{\frac{1}{12} \cdot \frac{24 \cdot 1.07^2 + 11 \cdot 1.13^2}{35}}$  [g/ml]

d   $101.12 \pm t_{35,0.975} \cdot \sqrt{\frac{1}{12} \cdot \frac{24 \cdot 1.07^2 + 11 \cdot 1.13^2}{35}}$  [g/ml]

e   $101.12 \pm t_{11,0.975} \cdot \sqrt{\frac{1}{11} \cdot \frac{24 \cdot 1.07^2 + 11 \cdot 1.13^2}{35}}$  [g/ml]

f  vet ej

2 I en multivariat regression får vi ett p-värde för F-teststorheten som är mindre än 0.01. Senare upptäcker vi i datorutskriften att en av variablernas t-teststorheter har ett p-värde större än 0.1. Vad kan man säga om detta.

- a  detta är inget förvånande
- b  det är fel på programvaran ty F-testet och t-testet mäter ungefär samma sak.
- c  om toleransvärdet är större än 0.1 så är detta normalt.
- d  om toleransvärdet är mindre än 0.1 så är detta normalt.
- e  inget av det ovanstående.
- f  vet ej.

3 Om vi låter  $r$  vara Pearson korrelationskoefficient kan vi tolka  $r^2$  i enkel linjär regression på följande sätt:

- a  om  $r^2$  är litet har vi en beskrivande modell
- b   $r^2$  beskriver den totala variationen
- c   $r^2$  mäter hur mycket som regressionen reducerar osäkerheten vid prediktion
- d   $r^2$  är summan av alla residualer och alltså ett mått på osäkerheten i modellen.
- e  inget av de ovanstående.
- f  vet ej.

4 I en intervjuundersökning av läsarna till tidskriften Fortune fann man att 46% hade andelar i aktiefonder, 63% ägde andelar i andra fondtyper och 74% ägde antingen den ena eller den andra eller båda typerna av fondandelar. Vilket av följande uttalande är då korrekt?

- a  sannolikheten att en läsare ägde båda typerna av fondandelar är 0.35 och sannolikheten att en läsare inte hade några fonder är 0.26
- b  sannolikheten att en läsare ägde båda typerna av fondandelar är 0.26 och sannolikheten att en läsare inte hade några fonder är 0.35
- c  sannolikheten att en läsare ägde båda typerna av fondandelar är 0.20 och sannolikheten att en läsare inte hade några fonder är 0.26
- d  sannolikheten att en läsare ägde båda typerna av fondandelar är 0.35 och sannolikheten att en läsare inte hade några fonder är 0.37
- e  inget av ovanstående
- f  vet ej

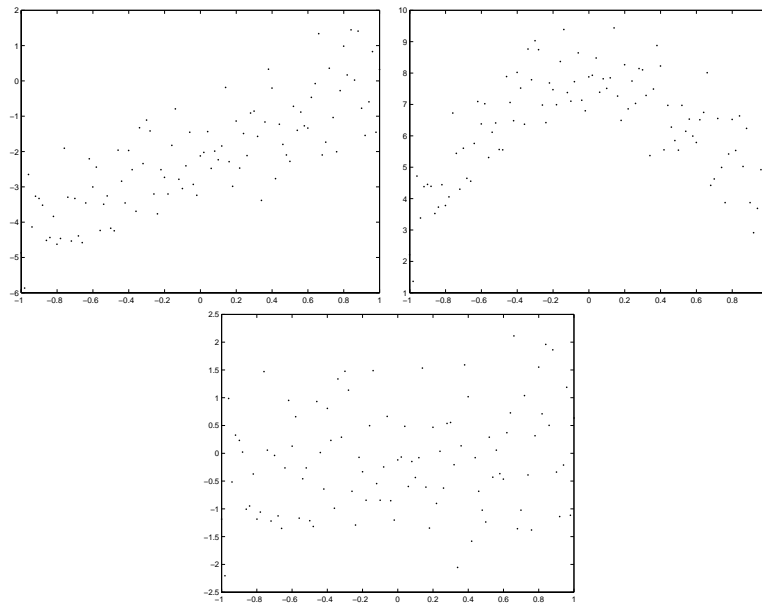
5 Som en del av en undersökning av matvaruaffärers placering av varor studerade man en grupp butiker som alla hade frukostvaror (havregryn, cornflakes, etc) på tre eller flera hyllor placerade ovanpå varandra.

30 butiker valdes ut slumpmässigt. Från 10 av butikerna valdes slumpmässigt en frukostvara från nedersta hyllan, från 10 en vara från den mellersta hyllan, och från 10 en vara från översta hyllan. För var och en av de 30 utvalda varorna bestämde man så sockerinnehållet, beräknat som gram per portion, från innehållsdeklarationen på paketet.

Med hjälp av dessa data vill man undersöka om det finns något sammanhang mellan hyllplacering och sockerinnehåll. Detta gör man bäst med

- a  regressionsanalys
- b  ensidig variansanalys utan blockindelning
- c  ensidig variansanalys med blockindelning (RCBD)
- d  Pearsons  $\chi^2$ -test in en tvåsidig tabell för kategoriska data
- e  inget av ovanstående
- f  vet inte

6 Nedstående bilder visar scatterplottar för observationer från tre olika bivariata fördelningar. Är det för någon/några av dem tydligt att korrelationskoefficienten är skilld från noll?



Figur 1: Observationer från fördelning a överst till vänster, fördelning b överst till höger och fördelning c nederst

- a  ja, för fördelning a och b, men inte för fördelning c
- b  ja, för fördelning a och c, men inte för fördelning b
- c  ja, men bara för fördelning c
- d  ja, för fördelning b och c, men inte för fördelning a
- e  nej, inte för någon av fördelningarna
- f  vet ej

7 Som del i tillverkning av elektroniska kretskort uppbygges en film genom oxidering av kiselskivor i en högtemperaturugn. Nedanstående tabell visar uppmätta värden av filmtjockleken (i ångström) och skivans avstånd (i cm) från ugnöppningen.

avstånd	10	15	20	25	30	35	40
tjocklek	725	1150	1250	1400	1350	1475	1600

För att undersöka filmtjocklekens beroende av avståndet från ugnöppningen gjordes en linjär regressionsanalys. Man fick då bl a följande datorutskrifter.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	406808.0357	406808.0357	25.84	0.0038
Error	5	78727.6786	15745.5357	.	.
C Total	6	485535.7143	.	.	.

Parameter Estimates					
Variable	DF	Estimate	Std Error	t Stat	Pr >  t
Intercept	1	675.8929	127.7023	5.29	0.0032
afstand	1	24.1071	4.7427	5.08	0.0038



Man vill bestämma ett 95% prediktionsintervall för filmtjockleken hos en skiva som ska placeras 30 cm från ugnöppningen. Detta ska räknas ut som:

$$\text{a } \square \quad 675.9 + 30 \cdot 24.11 \pm 2.5706 \cdot \sqrt{15745.5 \cdot \left( \frac{1}{7} + \frac{(30 - 25)^2}{700} \right)}$$

$$\text{b } \square \quad 675.9 + 30 \cdot 24.11 \pm 2.5706 \cdot \sqrt{15745.5 \cdot \left( 1 + \frac{1}{7} + \frac{(30 - 25)^2}{116.7} \right)}$$

$$\text{c } \square \quad 675.9 + 30 \cdot 24.11 \pm 2.5706 \cdot \sqrt{15745.5 \cdot \left( 1 + \frac{1}{7} + \frac{(30 - 25)^2}{700} \right)}$$

$$\text{d } \square \quad 675.9 + 30 \cdot 24.11 \pm 2.5706 \cdot \sqrt{15745.5 \cdot \left( 1 + \frac{30^2}{116.7} \right)}$$

$$\text{e } \square \quad 675.9 + 30 \cdot 24.11 \pm 2.5706 \cdot \sqrt{80922.6 \cdot \left( 1 + \frac{(30 \cdot 24.11)^2}{116.7} \right)}$$

f  vet ej

8 Efter injektion av ett antibiotikum i blodet binds en viss del av den injicerade mängden till serumproteiner. Detta fenomen har stor farmakologisk betydelse, eftersom det påverkar hur effektiv antibiotikan ifråga blir mot infektioner. I en studie ville man undersöka hur stor del av fem olika antibiotikatyper som bands. Varje medel injicerades på fyra olika individer. De tjugo frivilliga försökspersonerna tilldelades genom lottning en av de fem antibiotikatyperna.

Antibiotikum	M ängd bundet serum (okänd enhet)			
Penicillin G	29.6	24.3	28.5	32.0
Tetracycline	27.3	32.6	30.8	34.8
Streptomycin	5.8	6.2	11.0	8.3
Erythromycin	21.6	17.4	18.3	19.0
Chlomphenicol	29.2	32.8	25.0	24.2

Vilket av följande statistiska verktyg (med data på den form som ges i tabellen) skall du använda för att undersöka nollhypotesen att de fem typerna är likvärdiga?

- a   $\chi^2$ -test av radoberoende i en kontingenstabell
- b  enkel linjär regression
- c  ensidig variansanalys
- d  tvåsidig variansanalys
- e  vet ej

9 Vid en kontroll av antagandena i en enkel linjär regressionmodell är residualplottar ett mycket användbart verktyg. Man skall dock inte plotta residualerna mot:

a  de observerade värdena av svarsvariabeln.

b  de anpassade värdena av svarsvariabeln.

c  värdena av prediktorn.

d  värdena av regressorn.

e  tidsordningen för observationerna.

f  vet ej

10 Två uppmätta dimensioner,  $X$  och  $Y$  hos en plastdetalj tros följa en bivariat normalfördelning med  $\sigma_X = 0.04$ ,  $\sigma_Y = 0.08$ ,  $\mu_X = 3.0$ ,  $\mu_Y = 7.7$ ,  $\rho = 0$ . Då är  $P(2.95 < X < 3.05, 7.60 < Y < 7.80)$  lika med

- a  0.32
- b  0.42
- c  0.52
- d  0.62
- e  0.72
- f  Vet ej

- 11 Man var intresserad av felfrekvenser i ett sjukhuslaboratorium, och gick därför igenom 1000 slumpmässigt utvalda analysrapporter. Rapporterna kom från en vanlig arbetsvecka. De klassificerades som tillfredsställande, eller som behäftade med så pass viktiga fel att de borde ha gjorts om. Dessutom antecknades om analysen gjorts av dagskiftet eller kvällsskiftet.

Resultatet ges i nedanstående tabell:

Skift	felprocent	antal fel	antal OK	totalt	Procent
Dag	2.4	16	654	670	67
kväll	7.3	24	306	330	33
totalt	4.0	40	960	1000	100
Procent		4	96	100	

Pearson  $\chi^2$  testkvantitet för hypotesen om oberoende mellan felfrekvens og skift ges då av:

a   $\frac{(16 - 26.8)^2}{26.8} + \frac{(24 - 13.2)^2}{13.2}$

b   $\frac{(2.4 - 4.85)^2}{4.85} + \frac{(7.3 - 4.85)^2}{4.85}$

c   $\frac{(16 - 26.8)^2}{26.8} + \frac{(654 - 643.2)^2}{643.2} + \frac{(24 - 13.2)^2}{13.2} + \frac{(306 - 316.8)^2}{316.8}$

d   $\frac{(2.4 - 4.85)^2}{4.85} + \frac{(97.6 - 95.15)^2}{95.15} + \frac{(7.3 - 4.85)^2}{4.85} + \frac{(92.7 - 95.15)^2}{95.15}$

e   $\frac{(16 - 250)^2}{250} + \frac{(654 - 250)^2}{250} + \frac{(24 - 250)^2}{250} + \frac{(306 - 250)^2}{250}$

f  Vet ej

12 Ett företag som framställer insulin vill jämföra mätmetoderna för insulinstyrka vid 5 olika laboratorier runt om i världen.

För detta har man planerat göra följande sk “Round Robin” försök. Man skickar 6 prov från vart och ett av 3 insulinpartier till vart och ett av laboratorierna. Proverna analyseras sedan av laboratorierna, och analysresultaten skickas till företaget.

Resultaten (mätta i insulinstyrka per ml) kan antagas vara normalfördelade. Man vet av erfarenhet att insulinstyrkan kan variera lite från parti till parti.

Ändamålet med försöket är att undersöka om laboratorierna mäter samma insulinstyrka. Detta gör man bäst med

- a  regressionsanalys
- b  ensidig variansanalys utan blockindelning (one-way model)
- c  ensidig variansanalys med blockindelning (RCBD)
- d   $\chi^2$ -test för oberoende i en fyrfältsstabelle för kategoriska data
- e  inget av ovanstående
- f  vet ej

13 Den stokastiska variabeln  $X$  har väntevärde 3 och standardavvikelse 1 och variabeln  $Y$  har väntevärde 2 och standardavvikelse 2. Vidare är korrelationen mellan  $X$  och  $Y$  lika med 0.5. Sätt  $Z = X - Y + 2$ .

Då är variansen för  $Z$  lika med

- a  1
- b  2
- c  3
- d  5
- e  7
- f  vet ej

- 14 En teknolog gör examensarbete i en fabrik som tillverkar tandkräm. Hennes uppgift är att undersöka variationer i produktionen.

Tandkrämen fylls på tuber av en maskin med sex påfyllningskanyler som fyller parallellt från samma tank. Hon valde att bedömma variationen genom att undersöka ev. skillnader mellan tuber som kom från olika påfyllningskanyler. Under en dag med 17 timmars produktion tog hon en gång i timmen samtidigt ut en tub från varje kanyl. Hon vägde därefter tuberna. Från de uppmätta vikterna kunde hon så beräkna värdena i följande tabell:

Variation	SS
Tidpunkt	0.5458
Kanyl	4.0175
Error	0.1122

Testkvantiteten för test av hypotesen om att det inte finns någon skillnad i vikten av tandkrämstuber som kommer från olika påfyllningskanyler ges då av:

- a   $\frac{0.5458/16}{0.1122/80}$
- b   $\frac{0.5458/16}{0.1122/81}$
- c   $\frac{0.5458/5}{0.1122/81}$
- d   $\frac{0.5458/6}{0.1122/81}$
- e   $\frac{4.0175/5}{0.1122/80}$
- f  vet ej



15 En betongtillverkare vill jämföra dragstyrkan hos betong tillverkat med fyra olika produktionsmetoder. Vid tillverkningen används ofta fyra olika tillsättningsmedel, som kanske har inflytande på dragstyrkan. Man har därför planerat ett försök där var och en av produktionsmetoderna används tillsammans med vart och ett av de fyra tillsättningsmedeln. För att få tillräcklig precision i försöket tänker man göra tre upprepningar för varje kombination av metod och tillsättningsmedel.

Test av om produktionsmetoden har något inflytande på dragstyrkan görs bäst som:

- a  regressionsanalys
- b  ensidig variansanalys utan blockindelning (one-way model)
- c  ensidig variansanalys med blockindelning (RCBD)
- d   $\chi^2$ -test för oberoende i en fyrfältsstabell för kategoriska data
- e  inget av ovanstående
- f  vet ej