

TENTAMEN: Statistisk modellering för I3, TMS160, måndagen den 14 december 2003 kl ?? på V. **Jour:** John Gustavsson, ankn. 5316

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) 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 En fabrik som framställer penicillin vill jämföra 4 produktionsmetoder A, B, C och D.

Vid tillverkningen används ett majsderivat som råvara. Man vet att det är det finns kvalitetsskillnader mellan olika partier.

Man gjorde därför ett försök där man använde 5 olika partier majsderivat. För varje parti producerades penicillin med var och en av de 4 produktionsmetoderna. Utbytet av produktionen (penicillinhalten uttryckt i %) anges i nedanstående tabell.

| parti | metod | | | |
|-------|-------|----|----|----|
| | A | B | C | D |
| 1 | 89 | 88 | 97 | 94 |
| 2 | 84 | 77 | 92 | 79 |
| 3 | 81 | 87 | 87 | 85 |
| 4 | 87 | 92 | 89 | 84 |
| 5 | 79 | 81 | 80 | 88 |

Jämförelsen av produktionsmetoderna görs bäst med:

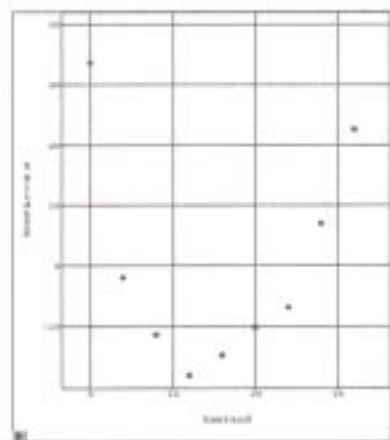
- a regressionsanalys
- b ensidig variansanalys utan blockindelning (one-way model)
- c ensidig variansanalys med blockindelning (RCBD)
- d χ^2 -test för oberoende i en tvåsidig tabell för kategoriska data
- e inget av ovanstående
- f vet inte

2 I en studie av däckslitage monterade man däck av ett visst fabrikat på en bil. För varje 4000 miles mätte man djupet av mönstret i tusendels tum på sex utvalda ställen på varje däck och registrerade medelvärdet av de så erhållna 24 mätningarna. Detta medelvärde kallades så slitbanedjup, y .

Man gjorde en linjär regressionsanalys av slitbanedjup, y , mot mätarställning i miles, x . Tabellen nedan visar de ursprungliga mätningarna, de anpassade värdena, \hat{y} og residualerna $e = y - \hat{y}$.

| Obs nr i | mätarställning x_i (miles) | Slitbanedjup y_i | anpassat värde \hat{y}_i | residual e_i |
|---------------|------------------------------------|-----------------------|-------------------------------|-------------------|
| 1 | 0 | 394.33 | 360.64 | 33.69 |
| 2 | 4000 | 329.50 | 331.51 | -2.01 |
| 3 | 8000 | 291.00 | 302.39 | -11.39 |
| 4 | 12000 | 255.17 | 273.27 | -18.10 |
| 5 | 16000 | 229.33 | 244.15 | -14.82 |
| 6 | 20000 | 204.83 | 215.02 | -10.19 |
| 7 | 24000 | 179.00 | 185.90 | -6.90 |
| 8 | 28000 | 163.83 | 156.78 | 7.05 |
| 9 | 32000 | 150.33 | 127.66 | 22.67 |

Residualerna är plottade mot mätarställning (i 1000 miles) i figuren nedan:



Residualerna tyder på att:

- a variansen beror på mätarställningen
- b mätningarna är inte oberoende
- c sambandet mellan slitage och körsträcka är inte linjärt
- d det är för lite data för att man ska kunna påvisa ett samband
- e det är låg förklaringsgrad (R^2), i analysen
- f vet inte

- 3 Som en del in en undersökning av resvanor gjordes ett slumpmässigt urval av personer som körde med bil till arbetet.

De utvalda personerna frågades om hur lång väg de körde till arbetet (3 svarskategorier), och om storleken av deras bil (4 svarskategorier) Antalet svar för var och en av de 12 möjliga svarskombinationerna finns i nedastående tabell

| bilstorlek | väglängd | | |
|------------|----------|----------|--------------|
| | 0-10 km | 10-20 km | ≥ 20 km |
| Mini | 6 | 27 | 19 |
| Compact | 8 | 36 | 17 |
| Standard | 21 | 45 | 33 |
| Caravan | 14 | 18 | 6 |

Dessa data analyseras 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 tvåsidig tabell för kategoriska data
- e inget av ovanstående
- f vet inte

4 En betongfabrik vill undersöka hur fuktupptagningen i den färdiggjutna betongen beror på dess sammansättning.

Man undersökte 5 olika recept för betongens sammansättning. För vart och ett av dessa gjöt man 6 betongprov som blev utsatta för fukt i 48 timmar. Därefter mätte man fukthalten i vart och ett av de 30 proven.

Eventuella skillnader i fukthalt för de 5 recepten undersöktes med en-sidig variansanalys utan blockindelning.

Modellantagandet om normalfördelning kan t ex undersökas med hjälp av en QQ-plot för:

- a de 30 uppmätta fukthalterna
- b modellens 30 anpassade värden av fukthalten
- c de 30 värdena av de standardiserade residualerna från den anpassade model för fukthalten
- d de 30 värdena av logaritmen av fukthalten
- e normalfördelningsantagandet kan inte undersökas med en QQ plot
- f vet inte

5 Resultatet av ett experiment där 400 gymnasie-elever i ett blindtest får välja mellan två sorters kakor är sammanfattat i följande tabell.

| | Sort A | Sort B | Totalt |
|------------------|--------|--------|--------|
| Manliga elever | 131 | 84 | 215 |
| Kvinnliga elever | 99 | 86 | 185 |
| Totalt | 230 | 170 | 400 |

Vid ett χ^2 -test av radandelar fås följande datorutskrift.

Expected counts are printed below observed counts

| | sort A | sort B | Total |
|-------|--------|--------|-------|
| 1 | 131 | 84 | 215 |
| | 123,63 | 91,38 | |
| 2 | 99 | 86 | 185 |
| | 106,38 | 78,63 | |
| Total | 230 | 170 | 400 |

$$\text{Chi-Sq} = 0,440 + 0,595 + 0,511 + 0,692 = 2,238$$

$$\text{DF} = 1, \text{ P-Value} = 0,135$$

Vad kan vi dra för slutsats från denna utskrift?

- a vi kan på signifikansnivå 0.05 säga att kvinnliga elever föredrar sort A framför B i större utsträckning än manliga elever.
- b vi kan på signifikansnivå 0.05 säga att manliga elever föredrar sort A framför B i större utsträckning än kvinnliga elever.
- c vi kan på signifikansnivå 0.05 säga att det inte är någon preferensskillnad mellan de manliga och kvinnliga eleverna.
- d vi kan på signifikansnivå 0.05 inte fastställa något preferensskillnad mellan de manliga och kvinnliga eleverna.
- e inget av ovanstående
- f vet inte.

6 I en undersökning av luftföroreningar i ett u-land tog man prover av luften vid fyra olika tidpunkter i perioden 1975-1976 och på fem ställen i landet. I varje prov bestämdes partikelinnehållet i mg/m^2 .

Resultaten anges i följande schema.

| tid- punkt | plats | | | | | medel- värde |
|-----------------|-------|-------|-------|-------|-------|-----------------|
| | 1 | 2 | 3 | 4 | 5 | |
| Okt 75 | 76 | 67 | 81 | 56 | 51 | 66.20 |
| Jan 76 | 82 | 69 | 96 | 59 | 70 | 75.20 |
| Maj 76 | 68 | 59 | 67 | 54 | 42 | 58.00 |
| Sep 76 | 63 | 56 | 64 | 58 | 37 | 55.60 |
| medel- värde | 72.25 | 62.75 | 77.00 | 56.75 | 50.00 | 63.75 |

Under antagande av en additiv effektmodell (dvs en modell utan växelverknings) finner man det anpassade värdet för plats 3 i maj 76 som:

- a 63.75
- b $(58.00 + 77.00)/2$
- c $58.00 + 77.00 - 63.75$
- d $67 - 63.75 + (58.00 + 77.00)/2$
- e $67 - 63.75 + \sqrt{58.00 \cdot 77.00}$
- f vet inte

- 7 En fabrik som tillverkar köttfärslimpor i en industriell ugn var intresserad av värmefördelningen i ugnen.

Man valde att undersöka värmefördelningen genom att mäta droppförlusterna från limporna i 8 positioner i ugnen. Man stekte 3 partier färs. Varje parti användes till 8 limpor. Limporna från de olika partierna placerades slumpmässigt på de 8 olika positionerna.

| Variation | SS |
|-----------|----------|
| position | 40.396 |
| parti | 1762.067 |
| error | 9.290 |
| total | 65.945 |

Den relevanta F-teststatistikan för test av hypotesen att det inte är skillnad på droppförlusterna i de olika positionerna är:

- a $(1762.067/2)/(9.290/14)$
- b $(1762.067/7)/(9.290/14)$
- c $(40.396/7) / (1762.067/2)$
- d $(40.396/7)/(9.290/7)$
- e $(40.396/7)/(9.290/14)$
- f vet inte

8 En forskare inom riskkapital gör en undersökning av hur information-flödet mellan entreprenören och investeraren fungerar. Vi får följande data:

- 51 investerare tycker det fungerar bra när entreprenören säger motsatsen.
- 38 investerare tycker det fungerar dåligt när entreprenören säger motsatsen.
- 41 investerare tycker det fungerar bra när entreprenören säger detsamma.
- 48 investerare tycker det fungerar dåligt när entreprenören säger detsamma.

dvs, vi har följande tabell:

| | <i>Entreprenören: "bra"</i> | <i>Entreprenören: "dåligt"</i> |
|------------------------|-----------------------------|--------------------------------|
| investeraren: "bra" | 41 | 51 |
| investeraren: "dåligt" | 38 | 48 |

Verkar det som om investeraren och entreprenören är överens om hur kommunikationen fungerar?

- a ja, för ett enkelt t-test ger ett p-värde större än 0.05.
- b ja, för ett χ^2 test ger ett p-värde mindre än 0.05
- c nej, för konfidensintervallet för $p_{nöjdentreprenör} - p_{nöjdinvesterare} - p_{missnöjdinvesterare}$ innehåller 0.
- d nej, för ett oberoendetest ger ett p-värde större än 0.05.
- e inget av de ovanstående svaren stämmer.
- f vet inte.

9 Statistiska Centralbyrån ställde på uppdrag av Fondbolagens Förening ett antal frågor om fondsparande inom ramen för SCB:s intervjuundersökning "Fråga Sverige" med ett slumpmässigt urval på 2 000 personer (18-74 år) fördelade över hela riket. Undersökningen genomfördes genom telefonintervjuer under september månad 1999. Svarsfrekvensen var 68 procent, vilket ansågs ge en god tillförlitlighet. Nedan följer några slutsatser.

Av alla svenskar sparar 49 procent genom direkt fondsparande. Motsvarande siffra för sparande genom fonder anknytna till pensionsparande är 28 procent. Av alla de svenskar som har fondsparande anknytna till pensionsparande har 62 procent också direkt fondsparande.

Andelen av alla svenskar som fondsparar är alltså:

- a 48 procent
- b 60 procent
- c 66 procent
- d 77 procent
- e inget av ovanstående
- f vet inte

- 10 En firma framställer plastdunkar i 6 olika gjutmaskiner. Man är intresserad av om medelvikten är den samma för alla 6 maskinerna.

För att undersöka detta tog man slumpmässigt ut 8 dunkar tillverkade i var och en av de 6 maskinerna och vägde dem (man vägde alltså totalt 48 dunkar).

Jämförelsen görs bäst med hjälp av:

- a jämförelse av konfidensintervall för väntevärdena av vikten för var och en av de 6 maskinerna
- b test i en ensidig variansanalysmodell utan blockindelning (one-way model)
- c test i en tvåsidig variansanalysmodell (two-way model)
- d Tukey's test för additivitet
- e χ^2 -test för oberoende i en tvåsidig tabell för kategoriska data
- f vet inte

- 11 En fabrik som framställer kullager med en nominell diameter av 1 micron använder två produktionslinjer. Man vill undersöka om det är någon skillnad mellan dessa.

Tabellen nedan visar diametrarna i 10 slumpmässigt uttagna kullager från var och en av produktionslinjerna.

| nr | Diameter | |
|----|----------|---------|
| | linje 1 | linje 2 |
| 1 | 1.18 | 1.72 |
| 2 | 1.42 | 1.62 |
| 3 | 0.69 | 1.69 |
| 4 | 0.88 | 0.79 |
| 5 | 1.62 | 1.79 |
| 6 | 1.09 | 0.77 |
| 7 | 1.53 | 1.44 |
| 8 | 1.02 | 1.29 |
| 9 | 1.19 | 1.96 |
| 10 | 1.32 | 0.99 |

Jämförelse av de två produktionslinjerna gör bäst med:

- a regressionsanalys
- b binomialfördelningstest
- c t-test for parade data
- d χ^2 -test for oberoende i en tvåsidig tabell for kategoriska data
- e ingen av ovanstående
- f vet inte

12 Tag två stokastiska variabler X och Y där korrelationen mellan de två variablerna är 0, dvs $\rho = 0$. Vi vet att $P[3.8 < X < 4.5] = 0.58$ samt $P[5.2 < Y < 8.9] = 0.12$. Vad kan vi då säga om $P[3.8 < X < 4.5, 5.2 < Y < 8.9]$?

- a $P[3.8 < X < 4.5, 5.2 < Y < 8.] = 0.7$
- b $P[3.8 < X < 4.5, 5.2 < Y < 8.] = 0.0696$
- c $P[3.8 < X < 4.5, 5.2 < Y < 8.] = 0.207$
- d $P[3.8 < X < 4.5, 5.2 < Y < 8.] = 0.46$
- e inget av ovanstående
- f vet inte

- 13 Ett företag vill undersöka om det är skillnad mellan felprocenterna i varor producerade i tre olika skift (dag, kväll, natt).

Man tog därför slumpmässiga stickprov av varor producerade i vart och ett av de tre skiften och räknade antalet defekta varor i varje stickprov. Resultatet finns i tabellen nedan.

| Klassifikation | Skift | | |
|----------------|-------|-------|------|
| | dag | kväll | natt |
| defekta | 45 | 55 | 70 |
| ikke-defekta | 905 | 890 | 870 |

Man gjorde ett χ^2 -test för oberoende i en tvåsidig tabell för kategoriska data, och fick teststorheten $\chi^2 = 6.29$.

Teststorheten ska jämföras med fraktilerna i en χ^2 -fördelning med:

- a 1 frihedsgrad
- b 2 frihedsgrader
- c 3 frihedsgrader
- d 4 frihedsgrader
- e inget av ovanstående
- f vet inte

14 Låt $p_1 = 0.1$ och $p_2 = 0.5$ vara felsannolikhet/år för 2 olika delsystem (A och B) som vart och ett kan stoppa en motor. Felen uppträder oberoende vilket innebär att mer än ett fel kan uppstå. Vad är sannolikheten för fel i delsystem A respektive delsystem B, givet att motorn stannat (och att inga andra orsaker till stoppet är möjliga)?

- a. 0.11, 0.77
- b. 0.17, 0.84
- c. 0.18, 0.91
- d. 0.32, 0.77
- e. vet inte.

- 15 För att utvärdera effekten av två brandimpregneringsmedel, A och B, gjordes följande försök. Fyra av åtta likadana polyestertygbitar behandlades med A och fyra med B. På samma sätt behandlades fyra av åtta likadana bomullstygbitar med A och 4 med B.

Tygbitarna utsattes så för eldflammar. Detta misslyckades emellertid för två av B-proverna, så att försöket kom att bli obalanserat.

Nedstående tabell visar resultaten, uttryckt som arean av den förbrända delen av tygbitarna.

| material | Impregnering | |
|-----------|--------------|-----------|
| | A | B |
| bomull | 39 | |
| | 36 | 28 |
| | 42 | 30 |
| | 45 | |
| polyester | 46 | 29 |
| | 50 | 25 |
| | 45 | height 32 |
| | 48 | 35 |

Dessa resultat blev analyserade med en obalanserad tvåsidig variansanalys med de två faktorerna *material* och *impregnering*. Tabellen nedan visar variansanalystabellerna för en typ I respektive typ III analys.

| Type I Tests | | | | | |
|--------------|----|----------------|-------------|--------|--------|
| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
| material | 1 | 14.8810 | 14.8810 | 1.28 | 0.2848 |
| impregnering | 1 | 730.1333 | 730.1333 | 62.67 | <.0001 |
| mat*impregn | 1 | 24.2000 | 24.2000 | 2.08 | 0.1801 |

| Type III Tests | | | | | |
|----------------|----|----------------|-------------|--------|--------|
| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
| material | 1 | 51.2000 | 51.2000 | 4.39 | 0.0625 |
| impregnering | 1 | 649.8000 | 649.8000 | 55.78 | <.0001 |
| mat*impregn | 1 | 24.2000 | 24.2000 | 2.08 | 0.1801 |

Kvadratsumman 730.1333 för impregnering i typ I analysen ger reduktionen i modellkvadratsumman

- a när man tar bort impregnering i en additiv modell som bara innehåller effekten impregnering
- b när man tar bort impregnering i en additiv modell som bara innehåller effekterna impregnering och material
- c när man tar bort material i en additiv modell som bara innehåller effekterna impregnering och material
- d är man tar bort impregnering i en additiv modell som innehåller effekterna impregnering och material samt växelverkan impregnering*material
- e inget av ovanstående
- f vet inte