– TAMS65 –

Problemsamling

Martin Singull

1. (222) En tillverkare av datorer har tagit emot ett mycket stort parti elektroniska komponenter, vars livslängd mätta i år skall vara oberoende och exponentialfördelade med väntevärdet 5. Man misstänker att säljaren blandat in en okänd andel a med komponenter, vars livslängder är oberoende och exponentialfördelade med vänte- värdet 1. Detta betyder att livslängden för en slumpmässigt vald komponent har täthetsfunktionen

$$f(x) = \frac{1}{5}e^{-x/5}(1-a) + e^{-x}a \quad \text{för } x > 0.$$

Man vill skatta a.

a) **Metod 1:** Man väljer slumpmässigt ut n_1 enheter ur partiet, och bestämmer deras livslängder $x_1, ..., x_{n_1}$. Skatta a med moment- metoden.
b) **Metod 2:** Man väljer slumpmässigt ut n_2 enheter ur partiet,
och använder dem under ett halvår och finner att y av dem har
gått sönder under detta halvår. Skatta a på lämpligt sätt.
c) Finns det någon praktisk fördel med metod 2?

2. (217) En viss typ av transistorer har exponentialfördelad livslängd. Man sätter 400 stycken i bruk samtidigt och konstaterar efter en tidsenhet att endast 109 stycken fungerar. Skatta medellivslängden och medianlivslängden.

3. (206) Låt $x_1, ..., x_n$ vara ett stickprov med täthetsfunktionen

$$f(x;\theta) = \begin{cases} \dfrac{1}{1-\theta} & \text{för } \theta \le x \le 1, \\ 0 & \text{för övrigt.} \end{cases}$$

Bestäm momentskattningen av θ och undersök om den är väntevärdesriktig.

4. Ytojämnheten har bestämts för fyra olika material som används för inkapsling . Resultat:

Material	Ytojämnhet						$\bar{x}_i$	s_i
EC10	0.50	0.55	0.55	0.36			0.4900	0.0898
EC10A	0.31	0.07	0.25	0.18	0.56	0.20	0.2617	0.1665
EC4	0.20	0.28	0.12				0.2000	0.0800
EC1	0.10	0.16					0.1300	0.0424

Gör parvisa jämförelser mellan materialen genom att konstruera lämpliga konfidensintervall, vart och ett med konfidensgraden 0.99. Du får anta att datamaterialet härrör från normalfördelning med samma varians. Man eftersträvar låg ytojämnhet.

5. För att jämföra effekten hos tre olika blodtryckssänkande mediciner behandlas tre grupper om vardera tio patienter med de olika medicinerna. Efter en månad mättes sänkningen av blodtrycket. Resultat:

	Medel-värde	Stickprovs-stand.avvik.
Medicin 1:	17.3	6.19
Medicin 2:	21.1	7.26
Medicin 3:	10.8	5.23

Modell: Vi har tre stickprov från $N(\mu_i, \sigma)$, $i = 1, 2, 3$.

a) konstruera ett konfidensintervall för σ av typen $(0, a)$ och med konfidensgraden 0.95.
b) Förefaller det troligt att $\mu_2 > 1.4\mu_3$? Besvara frågan genom att konstruera ett lämpligt konfidensintervall med konfidensgraden 0.95.

6. En liten industri har tre olika ugnar för upphettning av metallföremål. Man har förutsatt att ugnarna är inställda på samma temperatur men har börjat bli tveksam att så är fallet. Upprepade oberoende mätningar av temperaturen görs för varje ugn och man får följande värden:

						$\bar{x}_i$	s_i
Ugn 1:	492.4	493.6	498.5	488.6	494.0	493.42	3.55
Ugn 2:	488.5	485.3	482.0	479.4		483.80	3.96
Ugn 3:	502.1	496.8	497.5			498.80	2.88

Modell: Vi har tre stickprov från $N(\mu_i, \sigma)$, där μ_i är den inställda temperaturen för ugn nr i.

a) Konstruera ett uppåt begränsat 95% konfidensintervall för σ. b) Gör parvisa jämförelser mellan μ_i-na genom att konstruera lämpliga konfidensintervall, vart och ett med konfidensgraden 98 %.

7. I ett stickprov om 500 enheter från ett mycket stort parti befanns 87 vara felaktiga. Ange ett 95 % konfidensintervall för andelen felaktiga enheter.

8. Betjäningstiderna för ett kösystem är exponential- fördelade med väntevärde μ. Man har observerat 80 oberoende betjäningstider och fått medelvärdet $\bar{x} = 4.5$ minuter.

   a) Konstruera ett konfidensintervall för μ med konfidens- graden approximativt 0.95.

   b) Låt p vara sannolikheten att en betjäningstid är mer än tio minuter. Konstruera ett konfidensintervall för p med konfi- densgraden approximativt 0.95.

9. Ett företag har ett lager, där varor hämtas med truckar. Under 500 skilda, slumpmässigt valda tidsintervall av längden en timme noterades hur många truckar som anlände. Resultat:

Antal truckar:	0	1	2	3	4	5	6	7	8
Frekvens:	52	151	130	102	45	12	5	1	2

   Modell: $x_1, \ldots, x_{500}$ är observationer från $Po(\mu)$.

   Konstruera ett konfidensintervall för μ med konfidensgraden approximativt 0.95.

10. En halvledarfabrikant tillverkar enheter till en bilmotor. Kunden kräver att andelen felaktiga, p, högst är 0.05. Före leveransen av ett stort parti undersökte fabrikanten 200 enheter och fann bland dem fyra felaktiga.

    a) Pröva $H_0 : p = 0.05$ mot $H_1 : p < 0.05$ på nivån högst 0.05.

    b) Beräkna testets styrka för $p = 0.03$.

    c) Anser Du att kunden kan känna sig trygg efter resultatet i a)? Motivera ditt svar kortfattat.

    Lämpliga approximationer får utnyttjas i a) och b).

11. Med en viss mätutrustning registreras den radioaktiva bakgrundsstrålningen på en ort. Det är rimligt att anta att antalet registrerade partiklar under t minuter är $Po(\lambda t)$, där $\lambda = 5$ (enhet: min^{-1}). Efter ett radioaktivt utsläpp misstänker man att strålningen har ökat. Hur länge behöver man mäta strålningen, om

    $H_0 : \lambda = 5 \quad \text{mot} \quad H_1 : \lambda > 5$

    på nivån 0.01 med ett test som ger utslag med säkerheten 0.99, om strålningsintensiteten ökat med 50 %.

    Lämplig approximation får utnyttjas.

12. Vid 25 oberoende mätningar av en storhet har man erhållit

$$\bar{x} = 11.2 \quad \text{och} \quad s_x = 2.1$$

    och vi antar att $X_i \sim N(\mu, \sigma)$.

    a) Testa hypotesen $H_0 : \mu = 10$ mot $H_1 : \mu \neq 10$ på signifikans- nivån 1 %.

    b) Beräkna ett 99 % symmetriskt konfidensintervall för μ.

13. Draghållfastheter för tre sorters linor har bestämts. Resultat:

Typ	Uppmätta värden								$\bar{x}_i$	s_i
A:	3.72	2.93	4.73	3.90	4.57	4.27	5.38	3.24	4.09	0.809
B:	13.27	16.48	9.54	16.08	20.57	15.87	13.57	9.63	14.38	3.70
C:	10.42	11.98	11.50	7.85	5.71				9.49	2.65

    Kan vi säga vilken sort som är bäst?

a) Då vi skall besvara frågan kan vi inte använda en modell där vi betraktar data som tre stickprov från normalfördelningar med samma standardavvikelse. Visa detta med hjälp av lämpliga test vart och ett på nivån 0.01. Det räcker om du skriver ut ett av testen.

b) Antag i stället att de logaritmerade (ln) hållfastheterna utgör tre stickprov från $N(\mu_i, \sigma)$ och försök besvara frågan ovan genom att konstruera lämpliga konfidensintervall vart och ett med konfidensgrad 95%. Innan man genomförde försöket hade man ingen uppfattning om vilken typ som skulle vara bäst.

14. a) Livslängderna för 50 elektronrör bestämdes och man fick $\bar{x} = 38.5$. Man antar att livslängderna är oberoende och exponential- fördelade med ett okänt väntevärde μ. Konstruera ett tvåsidigt konfidensintervall för μ med konfidensgraden 95%.

    b) Då man såg hur observationerna fördelat sig på tallinjen, blev man tveksam över exponentialfördelningsantagandet:

Intervall	**Absolut frekvens**
$0 \leq x < 20$	14
$20 \leq x < 40$	18
$40 \leq x < 60$	7
$60 \leq x < 80$	6
$80 \leq x$	5

    Undersök med ett χ^2-test på nivån 0.10 om antagandet om exponentialfördelning är rimligt.

15. Låt X_1 och X_2 vara oberoende och $N(0,1)$ och låt den stokastiska vektorn $\boldsymbol{Y}$ definieras genom

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

    Bestäm täthetsfunktionen för $\boldsymbol{Y}$.

16. a) Vid en anställningsintervju får de arbetssökande genomgå tre olika test, som ger resultaten X_1, X_2 och X_3. Man har funnit att det för en slumpmässigt vald anställd i den aktuella branschen är rimligt att anta att

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 60 \\ 60 \\ 60 \end{pmatrix}, \begin{pmatrix} 100 & 80 & 20 \\ 80 & 100 & 10 \\ 20 & 10 & 80 \end{pmatrix} \right).$$

    För att få ett enklare beslutsunderlag sammanfattar man resultatet till ett värde $Y = (X_1 + X_2 + 2X_3)/4$. Bestäm fördelningen för Y samt ett värde a sådant att $P(Y > a) = 0.90$.

    b) Låt de s.v. X_1, X_2 och X_3 vara oberoende och $N(0,1)$. Sätt

$$U = X_1 - 2X_2 + X_3,$$
$$V = c_1X_1 + c_2X_2 + c_3X_3.$$

    Bestäm ett villkor på c_1, c_2 och c_3 som är nödvändigt och tillräckligt för att U och V ska vara oberoende.

17. I ett kommunikationssystem kan den i ett visst ögonblick mottagna signalen Y skrivas på formen $Y = X + Z$ där X är den verkliga utsända signalen och Z en störning som är oberoende av X. Vidare gäller att $X \sim N(10, 2)$ och $Z \sim N(0, 1)$.

    a) Bestäm fördelningen för den stokastiska vektorn med komponenter X och Z.

    b) Man vill rekonstruera X med hjälp av en linjär funktion $aY + b$ av den mottagna signalen. Bestäm konstanterna a och b så att $E(aY + b) = E(X)$ och $Var(X - aY - b)$ är minimal.

18. För en stokastisk följd $X_1, X_2, X_3, ...$ gäller att den stokastiska vektorn

$$\begin{pmatrix} X_{n-2} \\ X_{n-1} \\ X_n \end{pmatrix} \sim N\left(\begin{pmatrix} 20 \\ 20 \\ 20 \end{pmatrix}, \begin{pmatrix} 4 & -3.2 & 2.56 \\ -3.2 & 4 & -3.2 \\ 2.56 & -3.2 & 4 \end{pmatrix} \right)$$

med en kovariansstruktur som innebär att om man har två närliggande komponenter så är oftast den ena "stor" och den andra "liten".

a) Hur framgår detta av parametrarna i kovariansmatrisen?

b) För att få mindre slumpvariationer bildar man så kallade glidande medelvärden

$$Y_{n-1} = (X_{n-2} + X_{n-1})/2,$$
$$Y_n = (X_{n-1} + X_n)/2.$$

Bestäm fördelningen för den stokastiska vektorn med komponenterna Y_{n-1} och Y_n.

c) Beräkna korrelationen $\rho(Y_{n-1}, Y_n)$.

19. Låt X_1, X_2 och X_3 vara stokastiska variabler sådana att

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix} \right)$$

a) Beräkna $\mathrm{P}(X_2 > X_1 + 1)$.

b) Man vill göra en linjärkombination

$$Y = a_1 X_1 + a_2 X_2 + a_3 X_3$$

sådan att $\mathrm{E}(Y) = 5$ och $\mathrm{var}(Y)$ minimeras. Bestäm a_1, a_2 och a_3.

20. De stokastiska variablerna $X_1, X_2, \ldots, X_5$ är oberoende och $N(10, 2)$. Betrakta

$$Y_1 = \frac{1}{5}(X_1 + X_2 + \cdots + X_5),$$
$$Y_2 = 4X_1 + X_2 - X_3 - X_4 - X_5.$$

a) Bestäm simultana fördelningen för (Y_1, Y_2).

b) Beräkna $\mathrm{P}(Y_1 > Y_2)$

c) Beräkna korrelationskoefficienten mellan Y_1 och Y_2.

21. Låt X_1 och X_2 vara resultaten på två psykologiska test för en person och X_3 en betygssättning av personens förmåga att sköta en viss typ av arbetsuppgifter inom ett företag. Genom lång erfarenhet vet man att $\begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix}'$ har tredimensionell normalfördelning med väntevärdesvektor och kovariansmatris

$$\boldsymbol{\mu} = \begin{pmatrix} 71 \\ 53 \\ 18 \end{pmatrix} \quad ; \quad \boldsymbol{C} = \begin{pmatrix} 100 & 64 & 38 \\ 64 & 64 & 28.8 \\ 38 & 28.8 & 16 \end{pmatrix}.$$

Man vill ha information om X_3 via X_1 och X_2. Man kan bestämma en bästa linjär prediktor

$$\hat{X}_3 = a + bX_1 + cX_2$$

genom att kräva

(i) $\mathrm{E}(\hat{X}_3) = \mathrm{E}(X_3)$

(ii) $\mathrm{var}(X_3 - \hat{X}_3)$ minimal.

a) Bestäm denna bästa linjära prediktor $\hat{X}_3$.

b) Visa att predikteringsfelet $\epsilon = X_3 - \hat{X}_3$ är oberoende av X_1. (Predikteringsfelet är oberoende av X_2 också, men det behöver du inte visa.)

22. Den stokastiska variabeln $\begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix}'$ har en tredimensionell normalfördelning med väntevärdesvektor och kovariansmatris

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{respektive} \quad \begin{pmatrix} 7/2 & 1/2 & -1 \\ 1/2 & 1/2 & 0 \\ -1 & 0 & 1/2 \end{pmatrix}$$

a) Bestäm väntevärdesvektorn och kovariansmatris för $\begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix}'$, där

$$Y_1 = X_2 + X_3 \qquad Y_2 = X_1 + X_3 \qquad Y_3 = X_1 + X_2$$

b) Beräkna $\mathrm{P}(Y_2 > 2Y_3)$.

c) Undersök om Y_1 och Y_2 är oberoende. Svaret skall motiveras.

23. Störningarna $\varepsilon_1, \varepsilon_2, \varepsilon_3$ vid tre på varandra följande signalöverföringar i ett kommunikationssystem kan anses utgöra komponenterna i en normalfördelad vektor med väntevärdesvektor $\boldsymbol{\mu}$ och kovariansmatris $\boldsymbol{C}$ enligt nedan

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 1.5 & 0.9 & 0.54 \\ 0.9 & 1.5 & 0.9 \\ 0.54 & 0.9 & 1.5 \end{pmatrix}$$

Beräkna sannolikheten att medelvärdet

$$\bar{\varepsilon} = (\varepsilon_1 + \varepsilon_2 + \varepsilon_3)/3$$

av de tre störningarna till sitt belopp överstiger 2 enheter.

24. Visa att residualvektorn $\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ har kovariansmatrisen $(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\sigma^2$.

25. Betrakta den enkla linjära regressionsmodellen

$$Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j, \quad j = 1, 2, \ldots, n$$

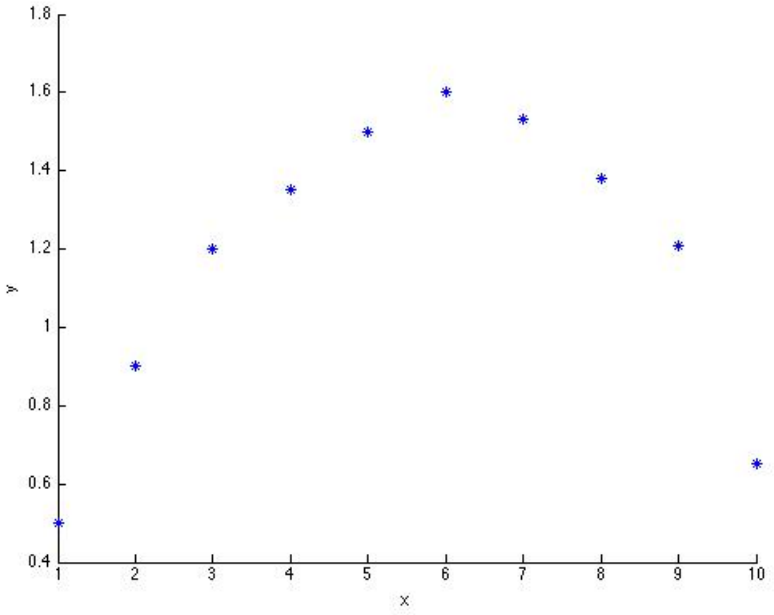
där $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ är oberoende och $N(0, \sigma)$.

Visa att minsta-kvadrat-metodens skattningar $\hat{\beta}_0$ och $\hat{\beta}_1$ är oberoende om och endast om

$$\sum_{j=1}^{n} x_j = 0.$$

26. En ny medicin mot cancer prövades på tio möss, som var och en hade en tumör av storleken 4 gram. Mössen fick olika doser (x) av medicinen, och minskningen (y) av tumören bestämdes för varje mus. Resultat:

x	1	2	3	4	5	6	7	8	9	10
y	0.50	0.90	1.20	1.35	1.50	1.60	1.53	1.38	1.21	0.65



Modell: $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ där $\varepsilon \sim N(0, \sigma)$.

En variansanalys enligt denna modell gav:

Variansanalys

Skattad regressionslinje: $y = -0.0283 + 0.551x - 0.0473x^2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-0.02833	0.07935
1	0.55150	0.03314
2	-0.047348	0.002936

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	1.26130
RES	7	0.03186
TOT	9	1.29316

a) Pröva på nivån 0.001

$H_0 : \beta_1 = \beta_2 = 0$ mot H_1 : minst en av β_1 och $\beta_2 \neq 0$.

b) Bestäm den dos som är optimal enligt den här regressionsanalysen.

c) Vore det tänkbart att ta bort en av förklaringsvariablerna? Motivera ditt svar kortfattat.

27. I en studie av lönsamheten för filmbolag har man valt ut 20 hollywoodfilmer slumpmässigt och för varje film tagit fram observerade värden på

y = bruttointäkt (enhet: miljoner dollar)

x_1 = produktionskostnad (enhet: miljoner dollar)

x_2 = marknadsföringskostnad (enhet: miljoner dollar).

Man var speciellt intresserad av om det hade betydelse om filmen baserats på en bok, som publicerats innan filmen producerades. För att separera sådana filmer från de övriga definierades en så kallad dummmy-variabel

$$x_3 = \begin{cases} 1 & \text{för film baserad på en bok} \\ 0 & \text{annars} \end{cases}$$

Resultat

	x_1	x_2	x_3	y
1	4.2	1.0	0	28
2	6.0	3.0	1	35
3	5.5	6.0	1	50
4	3.3	1.0	0	20
5	12.5	11.0	1	75
6	9.6	8.0	1	60
7	2.5	0.5	0	15
8	10.8	5.0	0	45
9	8.4	3.0	1	50
10	6.6	2.0	0	34
11	10.7	1.0	1	48
12	11.0	15.0	1	82
13	3.5	4.0	0	24
14	6.9	10.0	0	50
15	7.8	9.0	1	58
16	10.1	10.0	0	63
17	5.0	1.0	1	30
18	7.5	5.0	0	37
19	6.4	8.0	1	45
20	10.0	12.0	1	72

Data har analyserats enligt modellerna

Modell 1: $Y = \beta_0' + \beta_2' x_2 + \tilde{\varepsilon}$

Modell 2: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

där $\tilde{\varepsilon}$ respektive ε i de två modellerna antas vara oberoende och $N(0, \tilde{\sigma})$ respektive $N(0, \sigma)$. Variansanalyser finns nedan.

Variansanalys nr 1
Skattad regressionslinje: $y = 24.3 + 3.76x_2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	24.332	3.387
1	3.7606	0.4726

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	5094.6
RES	18	1448.3
TOT	19	6542.9

Variansanalys nr 2
Skattad regressionslinje: $y = 7.84 + 2.85x_1 + 2.28x_2 + 7.17x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	7.836	2.333
1	2.8477	0.3923
2	2.2782	0.2534
3	7.166	1.818

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	3	6325.2
RES	16	217.8
TOT	19	6542.9

a) Förklara kortfattat varför modell 2 beskriver datamaterialet bättre än modell 1. Motivera ditt svar med hjälp av lämpliga parametrar från datorutskriften.

b) Verkar det ha betydelse för intäkten om filmen baserats på en bok och i så fall på vilket sätt? Motivera ditt svar med hjälp av ett lämpligt 95% konfidensintervall.

c) En film baserad på en nyutkommen bok ska produceras. Man uppskattar produktionskostnaden till 11 miljoner dollar och man tänker satsa 9 miljoner dollar på marknadsföring av filmen. Uppskatta den förväntade bruttointäkten från filmen med hjälp av modell 2. Du behöver inte göra

något intervall.

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.399978 & -0.051254 & 0.006225 & -0.010691 \\ -0.051254 & 0.011308 & -0.004290 & -0.014215 \\ 0.006225 & -0.004290 & 0.004719 & -0.003020 \\ -0.010691 & -0.014215 & -0.003020 & 0.242792 \end{pmatrix}.$$

28. Antag att tiden Y för en kemisk reaktion har linjär regression med avseende på temperaturen x. Följande observerade data föreligger:

x_i	15	18	21	24	27
y_i	13.8	11.5	9.2	7.6	5.4

a) Beräkna punktskattningar av β_0, β_1 och σ^2.

b) Rita in värdena tillsammans med den beräknade regressionslinjen i ett koordinatsystem.

29. Vid ett test av bromsarna på en bil bromsades bilen upprepade gånger från en hastighet av cirka 100 km/h till stillastående på en torr asfaltväg. Vid varje försök mätte man dels hastigheten vid inbromsningens början (det var i praktiken svårt att hålla exakt hastigheten 100 km/h), dels bromssträckan. Resultat:

Begynnelsehastighet	103.5	98.0	95.5	102.0	100.0
Bromssträcka	57.0	51.5	50.0	56.0	54.0

a) Låt $Y_j, j = 1, 2, \ldots, 5$ beteckna bromssträckan i respektive försök. Antag att $Y_1, Y_2, \ldots, Y_5$ är oberoende och $N(\mu, \sigma)$, där μ betecknar den förväntade bromssträckan vid begynnelsehastigheten 100 km/h. (Vi betraktar sålunda all variation hos bromssträckan, även den som orsakas av varierande begynnelsehastighet, som ren slumpvariation.) Vi har alltså $Y_j = \mu + \varepsilon_j$ där $\varepsilon_j \sim N(0, \sigma)$. Bestäm ett 95% konfidensintervall för μ.

b) En del av variationen i bromssträckan beror troligen på att hastigheten inte varit exakt 100 km/h. Via en linjär modell bör man kunna ta hänsyn till hastigheten. I försök nummer j har vi haft hastigheten x_j och vi har fått bromssträckan y_j, som är observation av $Y_j = \beta_0 + \beta_1 x_j + \tilde{\varepsilon}_j$ där $\tilde{\varepsilon}_j \sim N(0, \tilde{\sigma})$. En variansanalys gav:

Variansanalys
Skattad regressionslinje: $y = -38.4 + 0.923x$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-38.423	6.169
1	0.92308	0.06179

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	34.338
RES	3	0.462
TOT	4	34.800

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 247.347 & \\ -2.47643 & 0.02481 \end{pmatrix}.$$

Beräkna ett 95% konfidensintervall för den förväntade bromssträckan vid hastigheten 100, dvs för $\beta_0 + 100\beta_1$.

c) Jämför resultaten i a) och b). Vilken analysmetod föredrar du?

30. Vid en fabrik tillverkas salpetersyra genom oxidation av ammoniak. Under 21 dagar har man mätt samhörande värden på

x_1 = luftflödet.
x_2 = ingående kylvattnets temperatur.
x_3 = koncentrationen av HNO_3 i den absorberande vätskan
y = 10× andelen (i %) av NH_3 som förloras, dvs ett omvänt måttpå utbytet.

Data from Operation of a Plant for the Oxidation of Ammonia to Nitric Acid.

Run No.	Air Flow x_1	Cooling Water Inlet Temp. x_2	Acid Concentration x_3	Stack Loss y
1	80	27	89	42
2	80	27	88	37
3	75	25	90	37
4	62	24	87	28
5	62	22	87	18
6	62	23	87	18
7	62	24	93	19
8	62	24	93	20
9	58	23	87	15
10	58	18	80	14
11	58	18	89	14
12	58	17	88	13
13	58	18	82	11
14	58	19	93	12
15	50	18	89	8
16	50	18	86	7
17	50	19	72	8
18	50	19	79	8
19	50	20	80	9
20	56	20	82	15
21	70	20	91	15

Variansanalyser enligt följande modeller finns nedan.

Modell 1: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

Modell 2: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

Konstruera med hjälp av modell 2 ett 95% konfidensintervall för E(Y) då $x_1 = 50$ och $x_2 = 18$. Är du nöjd med intervallet?

Variansanalys 1

Skattad regressionslinje: $y = -39.9 + 0.716x_1 + 1.30x_2 - 0.153x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-39.92	11.90
1	0.7156	0.1349
2	1.2953	0.3680
3	-0.1521	0.1563

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	3	1890.41
RES	17	178.83
TOT	20	2069.24

Variansanalys 2

Skattad regressionslinje: $y = -50.4 + 0.671x_1 + 1.30x_2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-50.359	5.138
1	0.6712	0.1267
2	1.2954	0.3675

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	1880.44
RES	18	188.80
TOT	20	2069.24

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 2.51724 & & \\ -0.01926 & 0.00153 & \\ -0.06189 & -0.00347 & 0.01288 \end{pmatrix}.$$

31. Man vill utnyttja en regressionsmodell för att beräkna priser på passagerarplan. Som beroende variabel har man

Y = flygplanets pris/antal passagerarplatser (enhet: 1 000-tal 8mm dollar)

och som förklaringsvariabler

x_1 = startvikten / antalet passagerarplatser

$x_2 = \ln$ (hastigheten)

Observerade värden:

x_1	x_2	y
249.3	5.44	172.00
272.3	5.59	194.44
219.6	5.65	190.00
213.7	5.50	129.55
216.8	5.59	148.91
190.6	5.66	135.16
226.8	5.56	116.07
233.9	5.66	166.67
220.6	6.12	150.00
222.4	6.12	177.57
225.7	5.61	178.57
236.0	5.50	115.39
199.9	5.59	154.41
252.6	5.95	198.86
224.1	5.95	181.37
212.9	5.36	127.78
211.1	6.14	169.23

Modell: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

Variansanalys

Skattad regressionslinje: $y = -283 + 0.688x_1 + 50.3x_2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-282.7	142.4
1	0.6881	0.2765
2	50.29	21.73

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	5119.3
RES	14	6605.7
TOT	16	11725.0

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 42.9585 & & \\ -0.04120 & 0.0001621 & \\ -5.8932 & 0.000825 & 1.0004 \end{pmatrix}.$$

Konstruera ett 95% prediktionsintervall för priset på ett flygplan med 60 platser, startvikten 15 000 och hastigheten 287. Är du nöjd med intervallet?

32. I mitten av 1800-talet ville den skotske fysikern James D. Forbes uppskatta höjden över havet genom att mäta kokpunkten för vatten. Han visste att höjden över havet kunde bestämmas med hjälp av lufttrycket. I en serie experiment studerade han sambandet mellan lufttryck och kokpunkt. Intresset för problemet motiverades av svårigheten för resenärer att transportera 1800-talets ömtåliga barometrar. Forbes samlade data i Alperna och Skottland, och följande datamaterial publicerades 1857:

Case No.	Boiling Point (^{o}F)	Pressure (in. Hg)	Log(Pressure)	100×Log(Pressure)
1	194.5	20.79	1.3179	131.79
2	194.3	20.79	1.3179	131.79
3	197.9	22.40	1.3502	135.02
4	198.4	22.67	1.3555	135.55
5	199.4	23.15	1.3646	136.46
6	199.9	23.35	1.3683	136.83
7	200.9	23.89	1.3782	137.82
8	201.1	23.99	1.3800	138.00
9	201.4	24.02	1.3806	138.06
10	201.3	24.01	1.3805	138.05
11	203.6	25.14	1.4004	140.04
12	204.6	26.57	1.4244	142.44
13	209.5	28.49	1.4547	145.47
14	208.6	27.76	1.4434	144.34
15	210.7	29.04	1.4630	146.30
16	211.9	29.88	1.4754	147.54
17	212.2	30.06	1.4780	147.80

En analys enligt

Modell 1: $Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j$ där $\varepsilon \sim N(0, \sigma)$, x_j = temperaturen och $y_j = 100 \cdot log$ (tryck), finns nedan. Residualplotten avslöjar en avvikande observation, en sk outlier, nämligen nr 12. Vi vill undersöka om det är troligt att avvikelsen har uppstått av en ren slump.

a) Datamaterialet har också analyserats enligt

Modell 2: $Y_j = \beta_0 + \beta_1 x_j + \beta_2 u_j + \varepsilon'_j$ där

$$u_j = \begin{cases} 1 & \text{för j=12} \\ 0 & \text{annars.} \end{cases}$$

Undersök med hjälp av ett lämpligt test eller konfidensintervall om observation nr 12 kan anses vara avvikande. Nivå 0.05.

b) I analys nr 3 har observation nr 12 strukits och återstående data har analyserats enligt modell 1. Konstruera utgående från denna analys ett 95% intervall med vars hjälp man kan undersöka om observation nr 12 kan anses vara avvikande. Vad blir din slutsats?

Anm. Man kan visa att metoderna i a) och b) är ekvivalenta, se Weisberg (1985), sid 115.

Variansanalys 1
Skattad regressionslinje: $y = -42.2 + 0.896x$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-42.164	3.341
1	0.89562	0.01646

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	425.76
RES	15	2.16
TOT	16	427.91

Variansanalys 2
Skattad regressionslinje: $y = -41.3 + 0.891x + 1.45u$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-41.335	1.003
1	0.891110	0.004944
2	1.4528	0.1174

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	427.73
RES	14	0.18
TOT	16	427.91

Variansanalys 3
Skattad regressionslinje: $y = -41.3 + 0.891x$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-41.335	1.003
1	0.891110	0.004944

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	419.19
RES	14	0.18
TOT	15	419.37

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 78.0093 & \\ -0.3843 & 0.001894 \end{pmatrix}.$$

33. En psykolog intresserade sig för effekterna av sömnbrist. 20 personer med likartade sömnvanor delades slumpmässigt in i fem grupper med fyra individer i varje. Ett försök genomfördes under två dagar. Natten mellan dag 1 och dag 2 fick:

    Grupp 1: 0 timmars sömn

    Grupp 2: 2 timmars sömn

    Grupp 3: 4 timmars sömn

    Grupp 4: 6 timmars sömn

    Grupp 5: 8 timmars sömn

    Dag 1: På morgonen fick varje person genomgå ett test, som gick ut på att utföra additioner av tal under tio minuter.

    Dag 2: Personerna fick på morgonen genomgå ett nytt men likvärdigt test med additioner av tal.

    För varje person beräknades skillnaden mellan antalet korrekt utförda additioner dag 1 och antalet korrekt utförda additioner dag 2. Resultat:

Grupp nr	Observerade data			
1	39	33	41	40
2	25	29	34	26
3	10	18	14	17
4	4	6	-1	9
5	-5	0	-3	-8

    Modell: För person nr i gäller att skillnaden i testresultat y_i är observation av $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ där $\varepsilon_1, \ldots, \varepsilon_{20}$ är oberoende och $N(0, \sigma)$ och där x_i = antalet timmar sömn.

    En variansanalys ges nedan.

    a) Konstruera ett 95% konfidensintervall för β_1 och pröva på nivån 0.05

    $$H_0 : \beta_1 = -4 \quad \text{mot} \quad H_1 : \beta_1 \neq -4.$$

    b) Undersök med hjälp av ett lämpligt test eller konfidensintervall om regressionslinjen skär x-axeln vid $x = 8$, dvs om de som sovit åtta timmar har oförändrad förmåga att klara testet. Nivå 5%.

Variansanalys
Skattad regressionslinje: $y = 38.1 - 5.43x$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	38.100	1.420
1	-5.4250	0.2898

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	4708.9
RES	18	241.9
TOT	19	4950.8

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.150000 & \\ -0.025000 & 0.006250 \end{pmatrix}.$$

34. Följande tabell visar utgifterna för privat konsumtion (y) samt den disponibla inkomsten (x_1) båda uttryckta i miljarder dollar. Variabeln x_2 anger krigstillståndet

$$x_2 = \begin{cases} 1 & \text{då landet är i krig} \\ 0 & \text{annars} \end{cases}$$

Uppgifterna gäller USA under åren 1935–1949.

x_1	x_2	y
58.5	0	56
66.3	0	62
71.2	0	67
65.5	0	64
70.3	0	67
75.7	0	71
92.7	0	81
116.9	1	89
133.5	1	99
146.3	1	108
150.2	1	120
160.0	0	144
169.8	0	162
189.2	0	175
188.6	0	178

Analys av datamaterialet enligt modellen

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

där $\varepsilon \sim N(0, \sigma)$ gav

Variansanalys
Skattad regressionslinje: $y = 1.00 + 0.92x_1 - 23.34x_2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	1.00157	2.38554
1	0.924056	0.0196041
2	-23.3432	2.06080

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	25868.1
RES	12	139.677
TOT	14	26007.777

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.488909 & & \\ -0.003625 & 0.000033 & \\ 0.006729 & -0.000889 & 0.364862 \end{pmatrix}.$$

a) På vilket sätt påverkas den privata konsumtionen av landets krigstillstånd enligt den här analysen? Motivera ditt svar genom att konstruera ett lämpligt tvåsidigt 99% konfidensintervall.

b) Konstruera ett 99% prediktionsintervall för den privata konsumtionen ett år, då $x_1 = 150$ och då landet inte är i krig.

35. Nedanstående material utgör data från 20 affärsmäns investeringar och motsvarande vinster. Om vi önskar investera kapitalet 20, vilken vinst kan vi då räkna med? Ta fram punktskattning och 95% intervallskattning.

Invest (x)	14	8	7	26	8	2	3	22	6	23
Vinst (y)	83	65	71	140	135	30	30	128	80	68
Invest (x)	29	4	13	14	7	5	13	6	5	8
Vinst (y)	139	88	121	125	56	98	101	96	73	116

Variansanalys
Skattad regressionslinje: $y = 57.5 + 3.55x$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	57.541	9.637
1	3.5524	0.0784

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	15327
RES	18	10972
TOT	19	26299

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.152361 & \\ -0.009180 & 0.000823 \end{pmatrix}.$$

36. Ett universitet har en dator som dels utnyttjas av lärare och elever, dels av externa användare. Låt

x_1 = antalet universitetsanvändare

x_2 = antalet externa användare

y = genomsnittlig svarstid i hundradels sekunder

Samhörande värden på x_1, x_2 och y har observerats:

x_1	x_2	y	x_1	x_2	y	x_1	x_2	y	x_1	x_2	y
10	0	8	48	4	52	23	3	13	69	12	106
36	8	59	21	1	13	66	7	81	58	10	74
75	5	77	66	10	88	10	2	15	26	2	23
16	4	21	30	3	28	70	10	100	14	4	18
35	5	45	55	9	70	44	0	28	62	9	72
50	8	56	49	6	63	42	4	43	63	3	55

En analys enligt modellen

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

har gjorts. Variansanalys finns nedan.

a) Pröva på nivån 0.001 $\quad H_0 : \beta_1 = \beta_2 = 0 \quad$ mot $\quad H_1$: minst en av β_1 och β_2 är $\neq 0$.

b) Belastar interna och externa användare datorn lika mycket, dvs är det möjligt att $\beta_1 = \beta_2$? Konstruera ett konfidensintervall för $\beta_1 - \beta_2$ med konfidensgraden 95%.

Variansanalys
Skattad regressionslinje: $y = -7.67 + 0.857x_1 + 3.90x_2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	-7.668	2.788
1	0.85692	0.08144
2	3.8957	0.4889

VARIANSANALYS

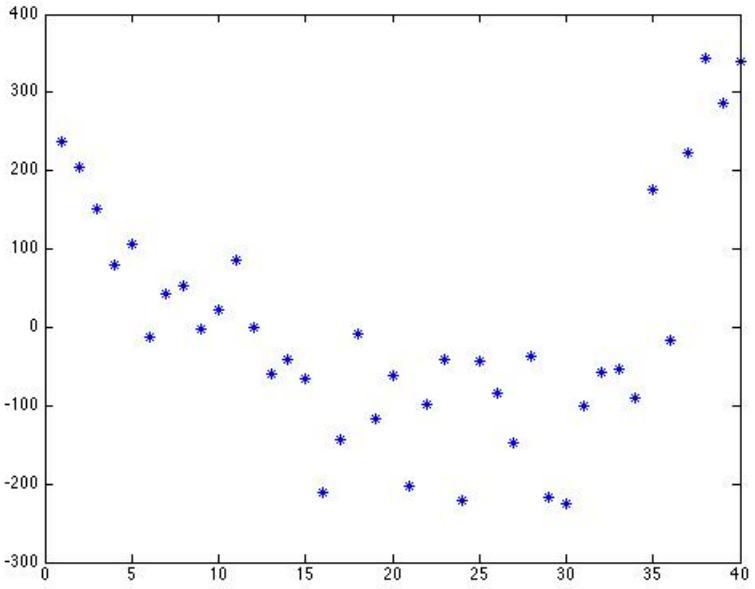
	Frihetsgrader	Kvadratsumma
REGR	2	19198.3
RES	21	707.0
TOT	23	19905.3

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.230838 & & \\ -0.004077 & 0.000197 & \\ -0.002392 & -0.000827 & 0.007098 \end{pmatrix}.$$

37. Man har observerat en tidsserie y_t, där $t = 1, 2, \ldots, 40$. För att göra en lämplig modell för Y_t gör man en analys enligt den preliminära modellen

$$Y_t = \beta_0 + \beta_1 t + \tilde{\varepsilon}_t$$

där $\tilde{\varepsilon}_t$-variablerna är oberoende och $N(0, \tilde{\sigma})$. Residualerna plottades mot t.



a) Efter att ha studerat residualplotten bestämmer man sig för att analysera datamaterialet enligt modellen

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

där ϵ_t-variablerna är oberoende och $N(0, \sigma\sqrt{t})$. Ange två skäl till att man väljer denna modell.

b) Ange en formel med vars hjälp $\beta_0, \beta_1, \beta_2$ i modellen i a) kan skattas. Eventuella matriser mm, som ingår i formeln, skall definieras fullständigt. (Ledning: Transformera datamaterialet på lämpligt sätt.)

38. För att jämföra tre olika sorters järnvägsräls lade man i fem olika distrikt ut två mil av vardera sorten och noterade under två år

x = genomsnittliga antalet tåg per dag, som passerade rälsavsnittet.

y = antalet sprickor i rälsen.

Resultat:

Typ A		Typ B		Typ C	
x	y	x	y	x	y
16.9	8	17.8	5	19.6	9
23.6	11	24.4	9	25.4	8
14.4	7	13.5	5	35.5	16
17.2	10	20.1	6	16.8	7
9.1	4	11.0	4	31.2	11

Modell: $Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 x + \varepsilon$ där $\varepsilon \sim N(0, \sigma)$ och där

$$z_1 = \begin{cases} 1 & \text{för typ B} \\ 0 & \text{för övrigt} \end{cases} \qquad z_2 = \begin{cases} 1 & \text{för typ C} \\ 0 & \text{för övrigt.} \end{cases}$$

En variansanalys finns nedan.

a) Verkar det genomsnittliga antalet tåg per dag vara av betydelse för sprickbildningen? Genomför ett lämpligt test på nivån 0.05 för modellen ovan.

b) Finns det skillnader mellan spårtyperna vad gäller sprickbildningen? Genomför ett lämpligt test på nivån 0.05.

Variansanalys
Skattad regressionslinje: $y = 1.39 - 2.66x_1 - 1.64x_2 + 0.41x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	1.39097	1.13679
1	-2.65579	0.823806
2	-1.64984	0.999121
3	0.406960	0.0601822

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	3	125.462
RES	11	18.5381
TOT	14	144.000

Variansanalys
Skattad regressionslinje: $y = 0.41 + 0.38x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	0.414181	1.29130
1	0.383768	0.0615918

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	107.877
RES	13	36.1277
TOT	14	144.000

39. Ett företag har jämfört efterfrågan av en viss sorts hygienartiklar i tolv olika försäljningsdistrikt. I fem av distrikten har varan sålts enbart genom detaljhandelskedjan A och i de sju övriga distrikten genom ett mycket stort antal detaljister. Följande resultat har erhållits:

Distrikt nr	**Distribution**	**Urbanise-ringsgrad**	**Relativ inkomst**	**Försäljnings-volym per inv**
1	Enbart A	42.2	31.9	167
2	-	48.6	33.2	185
3	-	42.6	28.7	170
4	-	39.0	26.1	152
5	-	34.7	30.1	150
6	Flera detaljister	44.5	28.5	192
7	-	39.1	24.3	183
8	-	40.1	28.6	180
9	-	45.9	20.4	191
10	-	36.2	24.1	171
11	-	39.3	30.0	168
12	-	46.1	34.3	189

Vi vill bland annat undersöka om distributionsformen kan ha betydelse för försäljningsvolymen genom att utnyttja en regressionsmodell.

Sätt

$$x_1 = \begin{cases} 1, & \text{om distribution genom enbart } A \\ 0, & \text{om distribution genom stort antal detaljister.} \end{cases}$$

x_2 = urbaniseringsgrad.

x_3 = relativ inkomst.

Y = försäljningsvolym per invånare.

Regressionsanalys med modellen $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ gav följande resultat:

Variansanalys
Skattad regressionslinje: $y = 83.8 - 16.2x_1 + 2.50x_2 - 0.208x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	83.84	16.87
1	-16.163	3.257
2	2.4953	0.3872
3	-0.2078	0.4313

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	3	2026.17
RES	8	211.49
TOT	11	2237.67

a) Verkar valet av distributionsform ha någon effekt?

b) Nedan följer några analyser för olika kombinationer av förklarande variabler. Vilken förklarande variabel skulle du välja, om du bara fick ta med en variabel?

Kan man vara säker på att det bästa paret av förklarande variabler innehåller den variabel du tog med ovan?

Hur många förklarande variabler (en, två eller tre) anser du rimligt att ta med i detta exempel?

Variansanalys
Skattad regressionslinje: $y = 182 - 17.2x_1$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	182.000	4.432
1	-17.200	6.866

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	862.9
RES	10	1374.8
TOT	11	2237.67

Variansanalys
Skattad regressionslinje: $y = 71.7 + 2.48x_2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	71.72	30.44
1	2.4833	0.7296

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	1201.0
RES	10	1036.6
TOT	11	2237.67

Variansanalys
Skattad regressionslinje: $y = 181 - 0.23x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	181.23	31.82
1	-0.226	1.112

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	9.2
RES	10	2228.5
TOT	11	2237.67

Variansanalys
Skattad regressionslinje: $y = 80.6 - 16.8x_1 + 2.44x_2$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	80.58	14.78
1	-16.761	2.880
2	2.4381	0.3524

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	2020.0
RES	9	217.6
TOT	11	2237.67

Variansanalys
Skattad regressionslinje: $y = 164 - 19.0x_1 + 0.645x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	164.48	26.56
1	-19.024	7.570
2	0.6449	0.9630

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	928.1
RES	9	1309.5
TOT	11	2237.67

Variansanalys
Skattad regressionslinje: $y = 89.4 + 2.76x_2 - 1.02x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	89.35	32.06
1	2.7572	0.7303
2	-1.0233	0.7591

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	1375.17
RES	9	862.50
TOT	11	2237.67

40. En viss typ av buss har bara en dörr, som måste användas både av dem som stiger av och dem som stiger på. Man har vid 20 tillfällen mätt tiden y (i sekunder) från det att bussen stannat vid en hållplats tills den åter satt igång. Samtidigt har man noterat antalet påstigande (x_1) och antalet avstigande (x_2).

x_1	x_2	y
0	1	4
2	3	24
1	0	23
1	0	12
2	1	20
4	0	45
5	8	60
1	1	18
0	1	5
1	0	15
1	3	18
8	3	88
5	0	50
1	3	24
1	1	12
0	3	8
0	6	14
1	4	16
2	0	32
0	8	25

Man vill försöka förklara bussens stopptid y med hjälp av en linjär regressionsmodell på x_1 och/eller x_2. En variansanalys ges nedan.

a) Om du bara fick använda en förklarande variabel i regressionen, vilken skulle du då välja? Motivera svaret. Gör den nytta som förklaringsvariabel? Genomför ett lämpligt test på nivån 1%.

b) Utgående från modellen i a) är det någon mening med att ta med även den andra förklarande variabeln i regressionen? Test skall genomföras på nivån 1%.

Skattade korrelationer ges av

	x_1	x_2
x_2	0.012	
y	0.960	0.192

Variansanalys
Skattad regressionslinje: $y = 8.74 + 9.40x_1$

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	7523.0
RES	18	635.5
TOT	19	8158.5

Variansanalys
Skattad regressionslinje: $y = 5.41 + 9.38x_1 + 1.46x_2$

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	7789.2
RES	17	369.3
TOT	19	8158.5

41. I en studie av överlevnadstiden för patienter med prostatacancer har man för varje patient noterat behandlingstypen (x_1), åldern i år (x_2) vid behandlingens början, halten (x_3) av ett visst karakteristiskt ämne, AP, i blodet, förekomsten av skelettmetastaser (x_4) samt överlevnadstiden (y) i månader räknat från behandlingens början. Vi har

$$x_1 = \begin{cases} 0 & \text{vid placebobehandling (= ingen behandling)} \\ 1 & \text{vid östrogenbehandling} \end{cases}$$

$$x_4 = \begin{cases} 0 & \text{om skelettmetastaser ej förekommer} \\ 1 & \text{om skelettmetastaser förekommer} \end{cases}$$

En variansanalys av datamaterialet för modellerna

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \tag{1}$$

där $\varepsilon \sim N(0, \sigma)$, och

$$Y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \tilde{\varepsilon} \tag{2}$$

där $\tilde{\epsilon} \sim N(0, \tilde{\sigma})$ gav de skattade korrelationerna

	x_1	x_2	x_3	x_4
x_2	-0.003			
x_3	0.008	-0.323		
x_4	0.206	-0.207	0.242	
y	0.182	-0.117	-0.344	-0.087

och

Variansanalys
Skattad regressionslinje: $y = 104.1 + 10.0x_1 - 0.96x_2 - 0.03x_3 - 4.32x_4$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	104.076	37.7264
1	10.0219	6.69623
2	-0.956706	0.506024
3	-0.0262767	0.00907681
4	-4.31777	7.18044

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	4	6665.0
RES	45	24082.1
TOT	49	30747.1

Variansanalys
Skattad regressionslinje: $y = 32.9 + 9.17x_1 - 0.022x_3$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	32.8860	4.97944
1	9.17159	6.65964
2	-0.0220795	0.00858002

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	2	4692.56
RES	47	26054.6
TOT	49	30747.1

a) Betrakta modell 1. Hur tycks östrogenhalten påverka överlevnadstiden? Motivera ditt svar genom att konstruera ett lämpligt konfidensintervall med konfidensgraden 80%, vilket egentligen är en väl låg konfidensgrad.

b) Variabeln x_3 fungerar bäst som ensam förklaringsvariabel (varför?), x_1 vill man ha med i modellen, eftersom ett av målen är att avgöra om östrogenbehandlingen gör nytta, men är det meningsfullt att ta med även x_2 och x_4? Genomför ett lämpligt test på nivån 5%.

c) Nämn någon fördel med regressionsanalyser av den typ vi gjort här jämfört med att dela in patienterna i två behandlingsgrupper och undersöka om det finns någon skillnad i förväntad överlevnadstid.

42. Följande datamaterial illustrerar tillväxten i en bakterieodling.

t	y
3	115 000
6	147 000
9	239 000
12	356 000
15	579 000
18	864 000

där t = antalet dagar efter inympningen och y = antalet bakterier.

Modell 1: $Y_t = e^{\beta_0+\beta_1 t+\varepsilon_t}$ där $\varepsilon_t \sim N(0,\sigma)$.

Konstruera ett 95% prediktionsintervall för antalet bakterier vid tidpunkten 20, dvs för Y_{20}.

Variansanalys
Skattad regressionslinje: $\ln(y) = 11.1 + 0.139t$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	11.1499	0.0619
1	0.138993	0.005300

VARIANSANALYS

	Frihetsgrader	Kvadratsumma
REGR	1	3.0428
RES	4	0.0177
TOT	5	3.0605

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.866667 & -0.066667 \\ -0.066667 & 0.006349 \end{pmatrix}.$$

Svar

1. a) $\hat{a} = (5 - \bar{x})/4$ (endast intressant då $0 < \hat{a} < 1$)
   b) $\hat{a} = \left(\frac{y}{n_2} - 0.0952\right)/0.2983$ (endast intressant då $0 < \hat{a} < 1$)

2. $x = 109$ är en observation av $X \sim Bin(400, p)$ där $p = e^{-1/\mu}$. Skatta p med $\hat{p} = 109/400$ ger $\hat{\mu} = \frac{1}{\ln(400/109)} \approx 0.77$.
   Medianlivslängden skattas med $\hat{\mu} \ln 2 \approx 0.53$.

3. $\hat{\theta} = 2\bar{x} - 1$ och $\mathrm{E}(\hat{\Theta}) = \theta$ d.v.s. skattningen är en väntevärdesriktig skattning av θ.

4. $I_{\mu_1-\mu_2} = (-0.03, 0.48)$, $I_{\mu_1-\mu_3} = (-0.01, 0.59)$, $I_{\mu_1-\mu_4} = (0.02, 0.70)$ osv. ($s = 0.1270$ med $df = 11$). Material 1 är sämre än material 4. Övriga skillnader är inte signifikanta.

5. a) $I_\sigma = (0, 8.12)$
   b) $I_{\mu_2 - 1.4\mu_3} = (0.17, \infty)$ dvs. preparat 2 är minst 40% bättre än preparat 3 med stor sannolikhet.

6. a) $I_\sigma = (0, 5.86)$
   b) $I_{\mu_1-\mu_2} = (2.89, 16.35)$; $I_{\mu_1-\mu_3} = (-12.71, 1.95)$; $I_{\mu_2-\mu_3} = (-22.67, -7.33)$.
   Ugn 2 tycks vara inställd på en lägre temperatur än ugn 1 och 3. Ingen signifikant skillnad mellan ugn 1 och 3.

7. X = antal felaktiga bland 500. X är Bi$(500, p)$. Approximation: $Bi \to N$; ger intervallet $I_p = (0.141, 0.207)$.

8. a) $I_\mu = \left(\frac{\hat{\mu}}{1+1.96/\sqrt{80}}, \frac{\hat{\mu}}{1-1.96/\sqrt{80}}\right) \simeq (3.69, 5.76)$
   b) $I_p = (0.067, 0.176)$

9. $\hat{\mu} = \bar{x} = 2.02$ med $\sum_{i=1}^{500} X_i \sim Po(500\mu)$ approximativt $N(500\mu, 500\mu)$ vilket ger $\hat{\mathrm{M}} = \bar{X} = \frac{1}{500}\sum_{i=1}^{500} X_i$ approximativt $N\left(\mu, \sqrt{\frac{\mu}{500}}\right)$. Hjälpvariabeln $\frac{\bar{X}-\mu}{\sqrt{\bar{X}/500}}$ approximativt $N(0,1)$ ger $I_\mu = \left(\hat{\mu} \mp 1.96\sqrt{\frac{\hat{\mu}}{500}}\right) = (2.02 \mp 0.12) \simeq (1.90, 2.14)$.

10. a) $x = 4$ observation från $Bin(200, p)$; H_0 förkastas; $\alpha = 0.0293$.
    b) Styrkan = 0.2851 (poissonapproximation).
    c) Ja, eftersom det är liten risk att ett dåligt parti ”slinker igenom” kontrollen.

11. $t = 21.5$ min; använd normalapproximation.

12. a) Teststorhet $u = (\bar{x} - 10)/(s_x/\sqrt{25}) = 2.86$. Den stokastiska variabeln $U \sim t(24)$ om H_0 är sann. Ur $t(24)$- tabellen får vi den kritiska gränsen 2.80; $|u| > 2.80$. Alltså kan H_0 förkastas på nivån 0.01.
    Notera att H_0 förkastas $\Leftrightarrow |u| > 2.80$
    $\Leftrightarrow u < -2.80$ eller $u > 2.80$
    $\Leftrightarrow \bar{x} < 10 - 2.80 s_x/5$ eller $\bar{x} > 10 + 2.80 s_x/5$
    $\Leftrightarrow 10 > \bar{x} + 2.80 s_x/5$ eller $10 < \bar{x} - 2.80 s_x/5$
    $\Leftrightarrow 10 \notin I_\mu$ (jfr b).
    b) Hjälpvariabeln $(\bar{X} - \mu)/(S_x/5) \sim t(24)$. Instängning ger $I_\mu = (\bar{x} - 2.80 s_x/5,\ \bar{x} + 2.80 s_x/5) = (10.02, 12.38)$; $10 \notin I_\mu$. Alltså kan H_0 förkastas.

13. a) $H_0 : \sigma_1 = \sigma_2$ mot $H_1 : \sigma_1 \neq \sigma_2$ prövas med hjälp av $v = \frac{s_2^2}{s_1^2}$; den stokastiska variabeln $V \sim F(7.7)$ om H_0 sann; H_0 förkastas om $v < 1/8.89$ eller $v > 8.89$. Här $v = 20.92 > 8.89$. H_0 förkastas.
    b) $I_{\mu_2-\mu_1} = (0.9755, 1.5117)$; $I_{\mu_2-\mu_3} = (0.1152, 0.7264)$; $I_{\mu_3-\mu_1} = (0.5172, 1.1284)$; vart och ett med konfidensgrad 0.95. Lina nr 2 är bäst. (Man måste bestämma vilka konfidens- intervall som skall beräknas, innan man har sett försöks- resultatet. Därför måste man göra tre tvåsidiga intervall.)

14. a) $I_\mu = (27.8, 49.2)$ (Normalapproximation kan göras på två sätt.)

    b) Efter sammanslagning till fyra klasser får vi $Q = 4.90 > 4.60$ där 4.60 ges i $\chi^2(2)$-tabellen. Exponentialfördelning förkastas på nivån 0.10.

15. $f_{\boldsymbol{Y}}(y_1, y_2) = \frac{1}{6\pi} e^{-(5y_1^2 - 2y_1y_2 + 2y_2^2)/18}$

16. a) $Y \sim N(60, \sqrt{50})$, $a = 50.9$

    b) Eftersom $(U, V)'$ är normalfördelad, är komponenterna oberoende om och endast om kovariansmatrisen är en diagonalmatris, dvs. om $c_1 - 2c_2 + c_3 = 0$.

17. a) $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\begin{pmatrix} 10 \\ 10 \end{pmatrix}, \begin{pmatrix} 4 & 4 \\ 4 & 5 \end{pmatrix})$

    b) $a = 4/5$ och $b = 2$

18. a) $\rho(X_{n-1}, X_n) = -0.8$. En stark negativ korrelation innebär att den ena variabeln är ”stor” när den andra är ”liten” och tvärtom.

    b) $\begin{pmatrix} Y_{n-1} \\ Y_n \end{pmatrix} \sim N(\begin{pmatrix} 20 \\ 20 \end{pmatrix}, \begin{pmatrix} 0.4 & 0.04 \\ 0.04 & 0.4 \end{pmatrix})$

    c) $\rho = 0.10$

19. a) $1 - \Phi(0.577) \simeq 0.28$

    b) $Y = (X_1 + X_3)/2$.

20. a) $\boldsymbol{Y} \sim N\left(\binom{10}{20}, \left(\begin{smallmatrix} 0.8 & 1.6 \\ 1.6 & 80 \end{smallmatrix}\right)\right)$;

    b) $1 - \Phi(1.135) \simeq 0.13$;

    c) $\rho = 0.2$.

21. $\hat{X}_3 = -10.45 + \frac{23}{90} X_1 + \frac{7}{36} X_2$

22. a) $\boldsymbol{Y} = \boldsymbol{AX} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \boldsymbol{X}$. Då är $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{A} \cdot E(\boldsymbol{X}) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ och $\boldsymbol{C}_Y = \boldsymbol{A}\boldsymbol{C}_X\boldsymbol{A}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 3 & 5 \end{pmatrix}$

    b) $P(Y_2 > 2Y_3) = P(Y_2 - 2Y_3 > 0)$; $\quad Y_2 - 2Y_3 = (0 \quad 1 \quad -2)\mathbf{Y}$;

    $$Y_2 - 2Y_3 \sim N\left(\begin{pmatrix} 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & -2 \end{pmatrix} \boldsymbol{C}_Y \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} \right) = N(1, \sqrt{10});$$

    $P(Y_2 - 2Y_3 > 0) = \Phi(1/\sqrt{10}) \simeq 0.63$.

    c) Y_1 och Y_2 är oberoende, eftersom de har simultan normalfördelning med $\mathrm{cov}(Y_1, Y_2) = 0$.

23. $\bar{\varepsilon} \sim N(0, 1.02)$; $\quad P(|\bar{\varepsilon}| > 2) = 0.0478$.

24. Skriv $\boldsymbol{e}$ som en linjär transformation av $\boldsymbol{Y}$-vektorn.

25. Bilda kovariansmatrisen för parameterskattningarna.

26. a) $v = 138.55 > 21.69$ (ur $F(2.7)$-tabell). H_0 förkastas.

    b) $x = 5.82$ ger maximum för det skattade regressionsuttrycket.

    c) Med hänsyn till kurvans utseende bör inte någon av förklaringsvariablerna uteslutas.

27. a) Modell 2 har förklaringsgrad $R^2 = 96.7\%$ vilket är klart bättre än $R^2 = 77.9\%$ för modell 1.

    b) $I_{\beta_3} = (7.166 \mp 2.12 \cdot 1.818) = (3.312,\ 11.020)$. Ett manus baserat på en bok ser ut att ge högre bruttointäkter.

    c) $\hat{\beta}_0 + 11\hat{\beta}_1 + 9\hat{\beta}_2 + \hat{\beta}_3 = 66.83$. Den förväntade bruttointäkten är cirka 67 miljoner dollar.

28. a) $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ där

$$\boldsymbol{X}' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 15 & 18 & \dots & 27 \end{pmatrix}; \quad \boldsymbol{y}' = \begin{pmatrix} 13.8 & 11.5 & \dots & 5.4 \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 5 & 105 \\ 105 & 2295 \end{pmatrix}^{-1} \begin{pmatrix} 47.5 \\ 935.4 \end{pmatrix} = \frac{1}{450} \begin{pmatrix} 2295 & -105 \\ -105 & 5 \end{pmatrix} \begin{pmatrix} 47.5 \\ 935.4 \end{pmatrix} = \begin{pmatrix} 23.99 \\ -0.6900 \end{pmatrix}$$

$\hat{\beta}_0 + \hat{\beta}_1 \cdot x_i$: 13.64 11.57 9.50 7.43 5.36

σ^2 skattas med $s^2 = \frac{1}{3}\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0.050$.

29. a) $I_\mu = (50.0, 57.4)$

b) $I_{\beta_0+100\beta_1} = (53.4, 54.4)$

c) Den linjära modellen beskriver datamaterialet bäst; detta styrks både av plotten och variansskattningarna.

30. $I_{\mu_0} = (\hat{\mu}_0 \mp t \cdot s \cdot \sqrt{0.1153}) \simeq (4.2, 8.8)$. Värdena i I_{μ_0} verkar låga, vilket bland annat kan bero på att sambandet mellan y och x_2 inte är linjärt.

31. $U_0 = 60Y_0$ där $Y_0 = \beta_0 + 250\beta_1 + 5.659\beta_2 + \varepsilon_0$;

$I_{Y_0} = (123.79, 224.03)$ ger $I_{u_0} \simeq (7\,400, 13\,400)$. Intervallet är så långt att det är tveksamt om man har nytta av det.

32. a) $I_{\beta_2} = (1.20, 1.70)$ dvs $0 \notin I_{\beta_2}$ och observation nr 12 tycks avvika från de övriga.

b) Med hjälp av analys 3 gör vi ett prediktionsintervall för Y_{12} : $I_{Y_{12}} = (140.74, 141.23)$ dvs $y_{12} = 142.44 \notin I_{Y_{12}}$ och observation nr 12 tycks avvika.

33. a) $I_{\beta_1} = (-5.4250 \mp t \cdot 0.2898) = (-6.03,\ -4.82)$ där $t = 2.10$ ges i $t(18)$-tabell. $-4 \notin I_{\beta_1}$. Alltså kan H_0 förkastas på nivån 0.05.

b) Vi konstruerar ett konfidensintervall för $\beta_0 + 8\beta_1$.

$\hat{\beta}_0 + 8\hat{\beta}_1 = -5.30$; den stokastiska variabeln $\hat{\beta}_0 + 8\hat{\beta}_1 \sim N(\beta_0 + 8\beta_1, 0.15\sigma^2)$ eftersom $\text{var}(\hat{\beta}_0 + 8\hat{\beta}_1) = \sigma^2 \cdot (1 \quad 8)(\mathbf{X}^T\mathbf{X})^{-1}\binom{1}{8} = 0.15\sigma^2$.

σ^2 skattas med $s^2 = Q_{res}/18 = 3.666^2$; frihetsgrader: 18

Hjälpvariabeln $\frac{\hat{\beta}_0 + 8\hat{\beta}_1 - (\beta_0 + 8\beta_1)}{s\sqrt{0.15}} \sim t(18)$ och den ger

$I_{\beta_0+8\beta_1} = (\hat{\beta}_0 + 8\hat{\beta} \mp 2.10s\sqrt{0.15}) \simeq (-8.28, -2.32)$.

$0 \notin I_{\beta_0+8\beta_1}$. Alltså kan vi förkasta hypotesen att regressionslinjen skär x-axeln vid $x = 8$. Eftersom intervallet bara innehåller negativa värden tycks åtta timmars sömn förbättra testresultatet, vilket kan innebära att utvilade personer har en inlärningseffekt.

34. a) $I_{\beta_2} = (-23.3432 \mp t \cdot 2.0608) \simeq (-29.6, -17.1)$. Krigstillstånd verkar ge lägre privat konsumtion.

b) Vi söker ett prediktionsintervall för $Y_0 = \beta_0 + 150\beta_1 + \varepsilon_0$;

$m_0 = \begin{pmatrix} 1 & 150 & 0 \end{pmatrix}\boldsymbol{\beta}$; $\hat{m}_0 = \begin{pmatrix} 1 & 150 & 0 \end{pmatrix}\hat{\boldsymbol{\beta}} = 139.61$; notera att $\hat{\beta}_0$ = interceptskattningen;

$$\text{var}(\hat{m}_0) = \begin{pmatrix} 1 & 150 & 0 \end{pmatrix}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\begin{pmatrix} 1 \\ 150 \\ 0 \end{pmatrix} = 0.1439\sigma^2$$

Den stokastiska variabeln $Y_0 - \hat{m}_0 \sim N(0, 1.1439\sigma^2)$.

σ^2 skattas med $s^2 = Q_{RES}/12 = 11.640$; $s = 3.411$; frihetsgrader: 12.

Hjälpvariabeln $(Y_0 - \hat{m}_0)/(s\sqrt{1.1439}) \sim t(12)$ den ger

$I_{Y_0} = (\hat{m}_0 \mp t \cdot s \cdot \sqrt{1.1439}) = (128.5, 150.7)$ där $t = 3.05$.

35. Prediktionsintervall för vinsten $I_{Y_0} = (74, 183)$

36. a) $v = 285.1 > 9.8$ (ur $F(2,21)$-tabell). H_0 förkastas.

b) $\beta_1 - \beta_2 = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \boldsymbol{\beta}$; den stokastiska variabeln

$\hat{\beta}_1 - \hat{\beta}_2 \sim N(\beta_1 - \beta_2, 0.008949\sigma^2)$; $I_{\beta_1-\beta_2} = (-4.18, -1.90)$; $\quad 0 \notin I_{\beta_1-\beta_2}$. De externa användarna tycks belasta datorn hårdare.

37. a) Sambandet mellan y_t och t ser ut att vara en krökt kurva av andragradstyp. var(ε_t) ser ut att växa då t växer.

b) Skatta $\boldsymbol{\beta}$ genom att utgå från $Z_t = Y_t/\sqrt{t}$ som får en regressionsmodell utan konstantterm. $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$ där $\boldsymbol{X}$ är koefficientmatrisen som hör ihop med $\boldsymbol{Z}$-vektorn.

38. a) $I_{\beta_3} = (0.27, 0.54)$; $\quad 0 \notin I_{\beta_3}$. Antalet tåg per dag tycks vara av betydelse.

b) $H_0 : \beta_1 = \beta_2 = 0$ prövas med hjälp av $v = 5.22 > 3.99$. H_0 förkastas. Det finns med stor sannolikhet skillnader mellan spårtyperna.

39. a) Ja, eftersom konfidensintervallet för β_1 inte innehåller nollan.

b) Fråga 1: x_2. Fråga 2: Nej. Fråga 3: Modellen med x_1 och x_2 som förklaringsvariabler ger minst variansskattning.

40. a) x_1 är den bästa ensamma förklaringsvariabeln, eftersom den är starkast korrelerad med y. Teststorheten $u = 213.069 > 8.29$, varför x_1 gör nytta som förklaringsvariabel.

b) Teststorheten $v = 12.25 > 8.4$. Ta med även x_2 i modellen.

41. a) $I_{\beta_1} = (10.0219 \pm t \cdot 6.69623) = (10.0219 \pm 8.7051) \simeq (1, 19)$ där $t = 1.30$ ges i $t(45)$-tabell. Östrogenbehandling tycks öka överlevnadstiden.

b) x_3 är starkast korrelerad med y och fungerar därför bäst som ensam förklaringsvariabel. Vi prövar modell 2 mot modell 1, dvs

$$H_0 : \beta_2 = \beta_4 = 0 \quad \text{mot} \quad H_1 : \quad \text{någon av} \quad \beta_2 \quad \text{och} \quad \beta_4 \neq 0$$

med ett F-test. Teststorhet

$$v = \frac{(26054.6 - 24082.1)/2}{24082.1/45} = 1.84$$

$F(2,45)$-tabell ger kritiska gränsen ~ 3.2. $1.84 < 3.2$. H_0 kan ej förkastas. Modell 2 duger.

c) Genom att ta med extra förklaringsvariabler utöver behandlingstypen kan man förklara en del av variationen, vilket ger mindre variansskattning, och då är det lättare att upptäcka behandlingseffekter. Vidare kan man hitta variabler som har betydelse för överlevnadstiden mm.

42. $I_{Z_0} = I_{\ln Y_0} = (13.6858, 14.1736)$ dvs $I_{Y_0} = (878000, 1431000)$.