5. SANNOLIKHETSMODELLER

5.1

$\Omega = \{$en prick, två prickar, tre prickar, fyra prickar, fem prickar, sex prickar$\}$. Det finns 6 utfall

5.2

a) $\Omega = \{(\text{du, du}), (\text{du, läraren}), (\text{läraren, du}), (\text{läraren, läraren})\}$, ändligt utfallsrum

b) $\Omega = \{$6-0, 6-1, 6-2, 6-3, 6-4, 7-5,..., 0-6, 1-6,...$\}$ oändligt utfallsrun

5.3

a) slumpmässigt, oändligt, diskret

b) slumpmässigt, ändligt, diskret

c) deterministiskt

d) slumpmässigt, oändligt, kontinuerligt eller diskret beroende på om man mäter med "oändigt stor precision" eller ej

e) slumpmässigt, oändligt, kontinuerligt eller diskret (om t.ex. ålser anges i hela år)

f) slumpmässigt, oändligt, kontinuerligt

5.4

Utfallsrummet i övning 5.1 och sannolikheter, t.ex. $P(\text{en prick}) = P(\text{två prickar}) = \ldots = P(\text{sex prickar}) = \frac{1}{6}$

5.5

Utfallsrummet i övning 5.2 och sannolikheter, t.ex. $P(\text{du, du}) = 0,70$, $P(\text{du, läraren}) = 0,15$, $P(\text{läraren, du}) = 0,10$ och $P(\text{läraren, lärarn}) = 0,05$

5.6

Utfallsrummet är

$\Omega = \{(FFF), (FFP), (FPF), (PFF), (FPP), (PFP), (PPF), (PPP)\}$.

Räkna antal utfall som stämmer med den sökta händelsen.

a) $\frac{1}{2}$

b) $\frac{7}{8}$

c) $\frac{3}{8}$

d) $\frac{3}{8}$

5.7

Skriv upp utfallsrummet t.ex. som i exempel 5.1 och figur 5.1. Räkna antal utfall som stämmer med den sökta händelsen.

a) $\frac{1}{4}$

b) $\frac{3}{4}$

c) $\frac{1}{2}$

6. BERÄKNING AV SANNOLIKHETER

6.1

Låt A vara händelsen att få en femma eller en sexa på den ena tärningen och B att få en femma eller en sexa på den andra tärninigen. Då är $P(A) = P(B) = \frac{2}{6}$ och $P(A \cap B) = \frac{4}{36}$. Vi söker $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{2}{6} + \frac{2}{6} - \frac{4}{36} = \frac{20}{36} = \frac{5}{9}$.

6.2

Låt A vara händelsen att det är dåligt språk och B att det är för mycket våld i ett slumpmässigt valt program. Vi vet att $P(A) = 0,42$, $P(B) = 0,27$ och $P(A \cap B) = 0,10$. Vi söker $P\left(\overline{A} \cap \overline{B}\right)$. Det är nu lätt att fylla i en fyrfältstabell

	A	$\overline{A}$	
B	0,10	0,17	0,27
$\overline{B}$	0,32	0,41	0,73
	0,42	0,58	1,00

och vi ser att $P\left(\overline{A} \cap \overline{B}\right) = 0,41$.

6.3

Låt A vara händelsen att laget vinner nästa match och B att det vinner den därpå följande matchen. Vi vet att $P(A) = 0,10$, $P(B) = 0,30$ och $P\left(\overline{A} \cap \overline{B}\right) = 0,65$. Vi söker $P\left(\left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right)\right)$. Det är nu lätt att fylla i en fyrfältstabell

	A	$\overline{A}$	
B	0,05	0,25	0,30
$\overline{B}$	0,05	0,65	0,70
	0,10	0,90	1,00

$\left(A \cap \overline{B}\right)$ och $\left(\overline{A} \cap B\right)$ är ömsesidigt uteslutande så att $P\left(\left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right)\right) = 0,05 + 0,25 = 0,30$.

6.4

Vi vet att $P(A) = 0,3$, $P(B) = 0,2$ och $P(A \mid B) = 0,5$. Därför blir $P(A \cap B) = P(B)\,P(A \mid B) = 0,2 \cdot 0,5 = 0,1$ och vi kan fylla i fyrfältstabellen

	A	$\overline{A}$	
B	0,1	0,1	0,2
$\overline{B}$	0,2	0,6	0,8
	0,3	0,7	1,0

6.5

Om fyrfältstabellen kompletteras får vi

	Får cancer	Får ej cancer	
Rökare	0,0054	0,1546	0,16
Ej rökare	0,0006	0,8394	0,84
	0,0060	0,9940	1,0

$P(\text{Får cancer} \mid \text{Rökare}) = \frac{P(\text{Får cancer} \cap \text{Rökare})}{P(\text{Rökare})} = \frac{0,0054}{0,16} = 0,03375$

$P(\text{Får cancer} \mid \text{Ej rökare}) = \frac{P(\text{Får cancer} \cap \text{Ej rökare})}{P(\text{Ej rökare})} = \frac{0,0006}{0,84} \approx 0,0007$, vilket är 47,25 gånger större än $P(\text{Får cancer} \mid \text{Rökare})$.

6.6

Låt A vara händelsen att en slumpmässigt vald olycka är svår och B att olyckan orsakades av en onykter förare. Vi vet att $P(A) = 0,20$, $P(B) = 0,12$ och $P(B \mid A) = 0,50$. Vi söker $P(A \mid B)$. Vi ser att $P(A \cap B) = P(A)\,P(B \mid A) = 0,20 \cdot 0,50 = 0,10$. Det är nu lätt att fylla i en fyrfältstabell

	A	$\overline{A}$	
B	0,10	0,02	0,12
$\overline{B}$	0,10	0,78	0,88
	0,20	0,80	1,00

och det följer att $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0,10}{0,12} = \frac{5}{6} \approx 0,8333$.

6.7

A och $\overline{A}$ är ömsesidigt uteslutande så att $P\left(A \cap \overline{A}\right) = 0$. Därför blir $P\left(\overline{A} \mid A\right) = \frac{P\left(A \cap \overline{A}\right)}{P(A)} = 0$.

6.8

Låt A vara händelsen att ett slumpmässigt valt parti accepteras och B att ett slumpmässigt parti är dåligt. Vi vet att $P\left(\overline{A}\right) = 0,15$, $P(B \mid A) = 0,06$ och $P(B) = 0,10$. Vi ser att $P(A) = 1 - P\left(\overline{A}\right) = 1 - 0,15 = 0,85$ och alltså är $P(A \cap B) = P(A)\,P(B \mid A) = 0,85 \cdot 0,06 = 0,051$. Fyrfältstabellen blir då

	A	$\overline{A}$	
B	0,051	0,049	0,10
$\overline{B}$	0,799	0,101	0,90
	0,85	0,15	1,00

och det följer att $P\left(A \mid \overline{B}\right) = \frac{P\left(A \cap \overline{B}\right)}{P\left(\overline{B}\right)} = \frac{0{,}799}{0{,}90} \approx 0,8878$.

6.9

a) Låt A vara händelsen att myntet är det vanliga myntet och B händelsen att myntet visar krona då det kastas. Vi vet att $P(A) = P\left(\overline{A}\right) = \frac{1}{2}$, $P(B \mid A) = \frac{1}{2}$ och att $P\left(B \mid \overline{A}\right) = 1$. Vi söker $P(A \mid B)$. Eftersom $P(A \cap B) = P(A)\,P(B \mid A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ och $P\left(\overline{A} \cap B\right) = P\left(\overline{A}\right) P\left(B \mid \overline{A}\right) = \frac{1}{2} \cdot 1 = \frac{1}{2}$ får vi fyrfältstabellen

	A	$\overline{A}$	
B	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
$\overline{B}$	$\frac{1}{4}$	0	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{2}$	1

Därmed kan vi beräkna $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}$.

b) På motsvarande sätt som i uppgift a) men låt nu B vara händelsen att myntet visar krona i båda kasten. Då är $P(B \mid A) = \frac{1}{4}$ och att $P\left(B \mid \overline{A}\right) = 1$ så att vi får $P(A \cap B) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$ och $P\left(\overline{A} \cap B\right) = \frac{1}{2} \cdot 1 = \frac{1}{2}$ med den resulterande fyrfältstabellen

	A	$\overline{A}$	
B	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{5}{8}$
$\overline{B}$	$\frac{3}{8}$	0	$\frac{3}{8}$
	$\frac{1}{2}$	$\frac{1}{2}$	1

dvs. $P(A \mid B) = \frac{1/8}{5/8} = \frac{1}{5}$.

6.10

Låt A vara händelsen att en slumpmässigt valt post är fel och B att rättningsrutinen anger att en slumpmässigt vald post är felaktig. Vi vet att $P(A) = 0,05$, $P(B \mid A) = 0,95$ och $P\left(B \mid \overline{A}\right) = 0,001$. Detta ger att $P(A \cap B) = P(A)P(B \mid A) = 0,05 \cdot 0,95 = 0,0475$ och $P\left(\overline{A} \cap B\right) = P\left(\overline{A}\right)P\left(B \mid \overline{A}\right) = (1 - 0,05) \cdot 0,001 = 0,00095$. Fyrfältstabellen blir då

	A	$\overline{A}$	
B	0,0475	0,00095	0,04845
$\overline{B}$	0,0025	0,94905	0,95155
	0,05	0,95	1,00

a) Vi söker $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0,0475}{0,04845} \approx 0,9804$.

b) Vi söker $P\left(A \mid \overline{B}\right) = \frac{P\left(A \cap \overline{B}\right)}{P\left(\overline{B}\right)} = \frac{0,0025}{0,95155} \approx 0,0026$.

6.11

Låt A vara händelsen att en åtalad är skyldig och B händelsen att det blir en fällande dom. Vi vet att $P(A) = 0,80$, $P(B \mid A) = 0,92$ och $P\left(B \mid \overline{A}\right) = 0,02$.

a) Vi söker $P(B)$. Vi ser att $(A \cap B) = P(A)P(B \mid A) = 0,80 \cdot 0,92 = 0,736$ och $\left(\overline{A} \cap B\right) = P\left(\overline{A}\right)P\left(B \mid \overline{A}\right) = 0,20 \cdot 0,02 = 0,004$. Detta ger oss fyrfältstabellen

	A	$\overline{A}$	
B	0,736	0,004	0,74
$\overline{B}$	0,064	0,196	0,26
	0,800	0,200	1,00

och det följer att $P(B) = 0,74$.

b) Vi söker $P\left(\overline{A} \mid B\right) = \frac{P\left(\overline{A} \cap B\right)}{P(B)} = \frac{0,004}{0,74} = 0,0054$.

6.12

Låt A vara händelsen att en slumpmässigt vald kund köper en köksinredning och B att kunden accepterar inbjudan till lunch. Vi vet att $P(A) = 0,05$, $P(B \mid A) = 0,50$ och $P\left(B \mid \overline{A}\right) = 0,10$. Detta ger att $P(A \cap B) = P(A)\,P(B \mid A) = 0,05 \cdot 0,50 = 0,025$ och $P\left(\overline{A} \cap B\right) = P\left(\overline{A}\right) P\left(B \mid \overline{A}\right) = (1 - 0,05) \cdot 0,10 = 0,095$. Fyrfältstabellen blir då

	A	$\overline{A}$	
B	0,025	0,095	0,12
$\overline{B}$	0,025	0,855	0,88
	0,05	0,95	1,00

Sannolikheten att en kund köper köksinredning bland de som tackat ja till lunch är $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0,025}{0,12} \approx 0,2083$.

Sannolikheten att en kund köper köksinredning bland de som tackat nej till lunch är $P\left(A \mid \overline{B}\right) = \frac{P\left(A \cap \overline{B}\right)}{P\left(\overline{B}\right)} = \frac{0,0025}{0,88} \approx 0,0284$.

6.13

Låt A vara händelsen att kursen för Statistiktjänst ökar och B händelsen att kursen för Inferenspraktik ökar. Vi vet att $P(B \mid A) = 0,8$, $P\left(\overline{B} \mid \overline{A}\right) = 0,8$ och $P(A) = 0,4$.

a) Vi söker $P(A \cap B) = P(A)\,P(B \mid A) = 0,4 \cdot 0,8 = 0,32$.

b) Vi söker $1 - P\left(\overline{A} \cap \overline{B}\right) = 1 - P\left(\overline{A}\right) P\left(\overline{B} \mid \overline{A}\right) = 1 - (1 - 0,4)\,0,8 = 1 - 0,48 = 0,52$.

6.14

Låt A vara händelsen att Cat Delicious blir en försäljningsframgång B att Active Juvenile blir en försäljningsframgång. Vi vet att $P(B \mid A) = 0,8$, $P\left(\overline{B} \mid \overline{A}\right) = 0,8$ och $P(A) = 0,4$. Detta ger att $P(A \cap B) = P(A)\,P(B \mid A) = 0,4 \cdot 0,8 = 0,32$ och $P\left(\overline{A} \cap \overline{B}\right) = P\left(\overline{A}\right) P\left(\overline{B} \mid \overline{A}\right) = (1 - 0,4) \cdot 0,8 = 0,48$. Fyrfältstabellen blir då

	A	$\overline{A}$	
B	0,32	0,12	0,44
$\overline{B}$	0,08	0,48	0,56
	0,40	0,60	1,00

a) $P(A \cap B) = 0,32$

b) $1 - P\left(\overline{A} \cap \overline{B}\right) = 1 - 0,48 = 0,52$ ($\overline{A} \cap \overline{B}$, "ingen produkt blir en försäljningsframgång", är komplementhändelse till händelsen "minst en produkt blir en försäljningsframgång).

6.15

Låt A vara händelsen att företaget läggs ner och B händelsen att företaget får ordern. Vi vet att $P(B) = 0,2$, $P(A \mid B) = 0,1$ och $P\left(A \mid \overline{B}\right) = 0,5$. Vi söker $P(A)$. Eftersom $P(A \cap B) = P(B)\,P(A \mid B) = 0,2 \cdot 0,1 = 0,02$ och $P\left(A \cap \overline{B}\right) = P\left(\overline{B}\right) P\left(A \mid \overline{B}\right) = 0,8 \cdot 0,5 = 0,40$ blir fyrfältstabellen

	A	$\overline{A}$	
B	0,02	0,18	0,2
$\overline{B}$	0,40	0,40	0,8
	0,42	0,58	1,0

dvs. $P(A) = 0,42$.

6.16

Låt A vara händelsen att ett slumpmässigt valt hus är av mexitegel och B att det har garage. Vi vet att $P(A \cap B) = 0,04$, $P\left(\overline{A} \cap \overline{B}\right) = 0,58$ och $P(B) = 0,4$. Detta ger att $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0,04}{0,4} = 0,10$.

6.17

Låt A vara händelsen att MOS får ordern och B händelsen att värdet på företagets aktier ökar. Vi vet att $P(A) = 0,2$, $P(B \mid A) = 0,8$ och $P\left(B \mid \overline{A}\right) = 0,1$. Vi söker $P(B)$. Eftersom $P(A \cap B) = P(A)\,P(B \mid A) = 0,2 \cdot 0,8 = 0,16$ och $P\left(\overline{A} \cap B\right) = P\left(\overline{A}\right) P\left(B \mid \overline{A}\right) = (1 - 0,2) \cdot 0,1 = 0,08$ får vi fyrfältstabellen

	A	$\overline{A}$	
B	0,16	0,08	0,24
$\overline{B}$	0,04	0,72	0,76
	0,20	0,80	1,0

och därmed är $P(B) = 0,24$.

6.18

Låt B vara händelsen att Hanna blir antagen till Beckmans konstskola och K att hon blir antagen till Konstfack. Vi vet att $P(B) = 0,7$ och $P(K) = 0,4$. Komplementhändelsen till "antagen till högst en skola" är "antagen till båda skolorna", dvs. $1 - P(A \cap B) = 0,75$. Detta ger fyrfältstabellen

	B	$\overline{B}$	
K	0,25	0,15	0,40
$\overline{K}$	0,45	0,15	0,60
	0,70	0,30	1,00

och sannolikheten att Hanna kommer in på minst en skola är 0,25+0,15+0,45=0,85.

6.19

Låt K vara händelsen att Håkan blir antagen till KTH och C händelsen att han blir antagen till Chalmers. Vi vet att $P(K) = 0,6$, $P(C) = 0,6$ och $P(K \cup C) = 0,8$. Vi söker $P\left(\left(K \cap \overline{C}\right) \cup \left(\overline{K} \cap C\right)\right)$. Eftersom $P(K \cap C) = P(K) + P(C) - P(K \cup C) = 0,6 + 0,6 - 0,8 = 0,4$ får vi fyrfältstabellen

	K	$\overline{K}$	
C	0,40	0,20	0,60
$\overline{C}$	0,20	0,20	0,40
	0,60	0,40	1,00

och tydligen är $P\left(\left(K \cap \overline{C}\right) \cup \left(\overline{K} \cap C\right)\right) = P\left(K \cap \overline{C}\right) + P\left(\overline{K} \cap C\right) = 0,20 + 0,20 = 0,40$.

6.20

a) $P(\text{exakt en pojke och två flickor}) = P((\text{PFF}) \cup (\text{FPF}) \cup (\text{FFP})) = P(\text{PFF}) + P(\text{FPF}) + P(\text{FFP}) = 0,5^3 + 0,5^3 + 0,5^3 = 0,375$

b) På grund av oberoende påverkas inte sannolikheten för flicka av könen hos äldre syskon, dvs, $P\left(\mathrm{F}\right)=0,5$

c) $P\left(\mathrm{FFFF}\right)=0,5^{4}=0,0625$

6.21

Vi vet att $P\left(A\right)>0$ och $P\left(B\right)>0$ vilket innebär att $P\left(A\right)P\left(B\right)>0$, men om A och B är ömsesidigt uteslutande så är $P\left(A\cap B\right)=0$, dvs. $P\left(A\right)P\left(B\right)\neq P\left(A\cap B\right)$ och alltså måste A och B vara stokastiskt beroende.

6.22

Från övning 6.5 vet vi att P (Får cancer | Rökare) $=0,03375$ och P (Får cancer | Ej rökare) $\approx 0,0007$, dvs, P (Får cancer | Rökare) $\neq P$ (Får cancer | Ej rökare) och alltså råder beroende.

6.23

Sannolikheten att Casiopeja eller Sweet Sue vinner loppet är $0,15+0,30=0,45$. Sannolikheten att någon av de andra hästarna vinner är därför $1-0,45=0,55$. Om Casiopeja och Sweet Sue stryks ur loppet blir således sannolikheten att Pollux vinner $\frac{0,20}{0,55}=\frac{4}{11}$.

6.24

Antag oberoende mellan matcherna.

a) P(Andersson vinner turneringen) = P((Andersson slår Bertilsson) och (Andersson slår Carlsson)) = 0,7·(1-0,9) = 0,07

b) P(Bertilsson vinner turneringen) = P((Bertilsson slår Andersson) och (Bertilsson slår Carlsson)) = (1-0,7)·0,8 = 0,24

c) P(Carlsson vinner turneringen) = P((Carlsson slår Anderssom) och (Carlsson slår Bertilsson)) = 0,9·(1-0,8) = 0,18 så att P(ingen vinner turneringen) = 1 - (0,07 + 0,24 + 0,18) = 0,51

6.25

Sannolikheten att en enhet inte deformeras i en deloperation är $1-0,03=0,97$. Eftersom det är stokastiskt oberoende mellan deloperationer är sannolikheten att en enhet inte blir deformerad i någon deloperation $0,97^{5}$. Sannolikheten att en enhet blir deformerad i minst en deloperation blir därför $1-0,97^{5}\approx 0,1413$.

6.26

Vi vet att $P\,(K_1) = 0,95$, $P\,(K_2) = 0,90$ och $P\,(K_1 \cap K_2) = 0,89$.

a) $P\,(K_1 \cup K_2) = P\,(K_1) + P\,(K_2) - P\,(K_1 \cap K_2) = 0,95 + 0,90 - 0,89 = 0,96$

b) $P\,(K_1) \cdot P\,(K_2) = 0,95 \cdot 0,90 = 0,855$ men $P\,(K_1 \cap K_2) = 0,89$ så att $P\,(K_1) \cdot P\,(K_2) \neq P\,(K_1 \cap K_2)$

c) Om oberoende gäller är $P\,(K_1 \cap K_2) = P\,(K_1) \cdot P\,(K_2) = 0,95 \cdot 0,90 = 0,855$ och vi får att $P\,(K_1 \cup K_2) = P\,(K_1) + P\,(K_2) - P\,(K_1 \cap K_2) = 0,95 + 0,90 - 0,855 = 0,995$

d) En fyrfältsrabell är

	K_1	$\overline{K}_1$	
K_2	0,89	0,01	0,90
$\overline{K}_2$	0,06	0,04	0,10
	0,95	0,05	1,00

så att $P\left(\overline{K}_1 \mid \overline{K}_2\right) = \frac{P\left(\overline{K}_1 \cap \overline{K}_2\right)}{P\left(\overline{K}_2\right)} = \frac{0,04}{0,10} = 0,4.$

6.27

a) Eftersom aktiernas värden är stokastiskt oberoende av varandra får vi att (i) $P\,(\text{båda aktiernas värden ökar}) = P\left(\text{Ö}\right)P\left(\text{Ö}\right) = 0,4 \cdot 0,4 = 0,16$. (ii) $P\,(\text{båda aktiernas värden minskar}) = P\,(M)\,P\,(M) = 0,4 \cdot 0,4 = 0,16$. (iii) $P\,(\text{värdet på exakt en aktie minskar}) = P((\text{den ena minskar och den andra minskar inte) eller (den ena minskar inte och den andra minskar})) =$

$$P\left(M \cap \left(O \cup \ddot{O}\right)\right) + P\left(\left(O \cup \ddot{O}\right) \cap M\right) = P\,(M)\,P\left(O \cup \ddot{O}\right) + P\left(O \cup \ddot{O}\right)P\,(M)$$
$$= 0,4\,(0,2 + 0,4) + (0,2 + 0,4)\,0,4 = 0,48$$

b) På samma sätt som i a) får vi (i) $P\,(\text{båda aktiernas värden ökar}) = 0,8 \cdot 0,8 = 0,64$ (ii) $P\,(\text{båda aktiernas värden minskar}) = 0,1 \cdot 0,1 = 0,01$. (iii) $P\,(\text{värdet på exakt en aktie minskar}) = 0,1\,(0,1 + 0,8) + (0,1 + 0,8)\,0,1 = 0,18$.

6.28

Låt A vara händelsen att aktie A ökar sitt värde en viss dag och B att aktie B ökade sitt värde dagen innan. Vi vet att $P\,(A) = P\,(B) = 0,5$ och att

$P\left(\overline{A}\cap\overline{B}\right)=0,3$. Vi söker $P\left(A\mid B\right)$. En fyrfältstabell ger

	A	$\overline{A}$	
B	0,3	0,2	0,5
$\overline{B}$	0,2	0,3	0,5
	0,5	0,5	1,00

så att $P\left(A\mid B\right)=\frac{P(A\cap B)}{P(B)}=\frac{0,3}{0,5}=0,6$

6.29

Låt A vara händelsen att en person har köpt produkten och B händelsen att en person har sett tv-reklamen. Undersökningen tyder på att $P\left(A\right)=0,21$, $P\left(B\right)=0,41$ och $P\left(A\cap B\right)=0,13$.

a) Vi söker $P\left(A\mid B\right)=\frac{P(A\cap B)}{P(B)}=\frac{0,13}{0,41}\approx 0,3171$.

b) tv-reklamen har varit framgångsrik om andelen som köpt produkten är större bland de som sett reklamen än bland de som inte sett den, dvs. vi vill jämföra $P\left(A\mid B\right)$ med $P\left(A\mid\overline{B}\right)$. Vi får fyrfältstabellen

	A	$\overline{A}$	
B	0,13	0,28	0,41
$\overline{B}$	0,08	0,51	0,59
	0,21	0,79	1,00

och alltså blir $P\left(A\mid\overline{B}\right)=\frac{P\left(A\cap\overline{B}\right)}{P\left(\overline{B}\right)}=\frac{0,08}{0,59}\approx 0,1356$. Eftersom $P\left(A\mid B\right)>P\left(A\mid\overline{B}\right)$ drar vi slutsatsen att tv-reklamen har varit framgångsrik.

6.30

Låt A, B och C beteckna händelserna att en slumpmässigt vald komponent kommer från tillverkare A, B respektive C. Låt vidare F beteckna händelsen att en slumpmässigt vald komponent är felaktig. Vi vet att $P\left(A\right)=0,10$ och $P\left(B\right)=0,20$ så att $P\left(C\right)=1-\left(0,10+0,20\right)=0,70$. Vidare vet vi att $P\left(F\mid A\right)=0,03$, $P\left(F\mid B\right)=0,02$ och $P\left(F\mid C\right)=0,005$. Detta innebär att $P\left(F\cap A\right)=0,03\cdot 0,10=0,003$, $P\left(F\cap B\right)=0,02\cdot 0,20=0,004$ och $P\left(F\cap C\right)=0,005\cdot 0,70=0,0035$. Den sökta sannolikheten blir därför $P\left(F\right)=P\left(\left(F\cap A\right)\cup\left(F\cap B\right)\cup\left(F\cap C\right)\right)=0,003+0,004+0,0035=0,0105$

7. STOKASTISKA VARIABLER

7.1

$E(X) = 0 \cdot 0,3 + 1 \cdot 0,7 = 0,7$

$V(X) = (0 - 0,7)^2 \cdot 0,3 + (1 - 0,7)^2 \cdot 0,7 = 0,49 \cdot 0,3 + 0,09 \cdot 0,7 = 0,21$

7.2

x	$f(x)$	$xf(x)$	$x^2 f(x)$
0	1/8	0	0
1	3/8	3/8	3/8
2	3/8	6/8	12/8
3	1/8	3/8	9/8
Summa	1	12/8	24/8

$E(X) = \frac{12}{8} = 1,5$

$V(X) = E(X^2) - (E(X))^2 = \frac{24}{8} - \left(\frac{12}{8}\right)^2 = 0,75$

7.3

a) $P(A \text{ inträffar minst en gång}) = 1 - P(B \text{ inträffar alltid}) = 1 - \left(\frac{2}{3}\right)^4 = \frac{65}{81}$

b) $P(A \text{ inträffar minst två gånger}) = 1 - P(A \text{ inträffar färre än två gånger}) = 1 - (P(A \text{ inträffar aldrig}) + P(A \text{ inträffar en gång})) = 1 - (P(BBBB) + P(ABBB) + P(BABB) + P(BBAB) + P(BBBA)) = 1 - (\left(\frac{2}{3}\right)^4 + 4\frac{1}{3}\left(\frac{2}{3}\right)^3) = \frac{11}{27}$

c) $P(X = 0) = P(\text{B inträffar}) = \frac{2}{3}$, $P(X = 1) = P(\text{A inträffar}) = \frac{1}{3}$. $E(X) = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$, $V(X) = \left(0 - \frac{1}{3}\right)^2 \cdot \frac{2}{3} + \left(1 - \frac{1}{3}\right)^2 \cdot \frac{1}{3} = \frac{2}{9}$

d)

y	$f(y)$	$yf(y)$	$y^2 f(y)$
0	16/81	0	0
1	32/81	32/81	32/81
2	24/81	48/81	96/81
3	8/81	24/81	72/8
4	1/81	4/81	16/81
Summa	1	108/81	216/81

e) $E(Y) = \frac{108}{81} = \frac{4}{3}$, $V(Y) = E(Y^2) - (E(Y))^2 = \frac{216}{81} - \left(\frac{4}{3}\right)^2 = \frac{8}{9}$.

7.4

Vi vet att $E(X) = 7$. Bilda nu $Y = X - 7$. Detta ger tabellen

y	$f(y)$	$yf(y)$	$y^2 f(y)$
-5	1/36	-5/36	25/36
-4	2/36	-8/36	32/36
-3	3/36	-9/36	27/36
-2	4/36	-8/36	16/36
-1	5/36	-5/36	5/36
0	6/36	0	0
1	5/36	5/36	5/36
2	4/36	8/36	16/36
3	3/36	9/36	27/36
4	2/36	8/36	32/32
5	1/36	5/36	25/36
Summa	1	0	210/36

$E(Y) = 0$

$V(Y) = E(Y^2) - (E(Y))^2 = \frac{210}{36} - 0^2 \approx 5,83$

b) Bilda nu $Z = \frac{Y}{\sqrt{210/36}}$. Detta ger tabellen

z	$f(z)$	$zf(z)$	$z^2f(z)$
-2,07	1/36	-2,07/36	4,29/36
-1,66	2/36	-3,31/36	5,49/36
-1,24	3/36	-3,73/36	4,63/36
-0,83	4/36	-3,31/36	2,74/36
-0,41	5/36	-2,07/36	0,86/36
0	6/36	0	0
0,41	5/36	2,07/36	0,86/36
0,83	4/36	3,31/36	2,74/36
1,24	3/36	3,73/36	4,63/36
1,66	2/36	3,31/36	5,49/36
2,07	1/36	2,07/36	4,29/36
Summa	1	0	1

$E(Z) = 0$

$V(Z) = E(Z^2) - (E(Z))^2 = 1 - 0^2 = 1$

7.5

$E(Z) = E\left(\frac{X-E(X)}{\sqrt{V(X)}}\right) = \frac{1}{\sqrt{V(X)}}E(X - E(X)) = \frac{1}{\sqrt{V(X)}}(E(X) - E(X)) = 0$

$V(Z) = V\left(\frac{X-E(X)}{\sqrt{V(X)}}\right) = \frac{1}{V(X)}V(X - E(X)) = \frac{1}{V(X)}V(X) = 1.$

7.6

Låt X vara vinsten från arbetet.

a) $E(X) = 180000 \cdot 0,8 + (-60000) \cdot 0,2 = 132000$, ja, man kan förvänta sig en positiv vinst.

b) Låt p vara sannolikheten att arbetet blir färdigt inom kontraktstiden.

$E(X) = 180000 \cdot p + (-60000) \cdot (1 - p) = 0$

$240000p = 60000$

$p = 0,25$

7.7

Notera först att herr Johansson inte gör någon vinst med sannolikheten $1 - \left(\frac{3}{18} + \frac{7}{18} + \frac{3}{18}\right) = \frac{5}{18}$

Förväntad vinst är $-50000 \cdot \frac{3}{18} + 0 \cdot \frac{5}{18} + 50000 \cdot \frac{7}{18} + 100000 \cdot \frac{3}{18} = \frac{250000}{9} \approx 27778$ kr.

7.8

a)

x	$f(x)$
-496000	0,005
4000	0,995

b) $(-496000) \cdot 0,005 + 4000 \cdot 0,995 = 1500$

c) Antag att premien är x kr. Då är

$-(500000 - x) \cdot 0,005 + x \cdot 0,995 = 10000$

$x = 12500$ kr

7.9

a) Förväntad vinst vid spel på Aldebaran $= 250 \cdot 0,2 - 100 = -50$ kr

Förväntad vinst vid spel på Blossom Star $= 120 \cdot 0,4 - 100 = -52$ kr

Förväntad vinst vid spel på Cassiopeja $= 400 \cdot 0,1 - 100 = -60$ kr

b) Oavsett vilken han spelar på får han en förväntad förlust. Han bör alltså inte spela på någon.

7.10

a) $P(X > 5,7) = 1 - P(X \leq 5,7) = 1 - F(5,7) = 1 - (5,7 - 5)^2 (3 \cdot 5,7^2 - 38 \cdot 5,7 + 121) = 1 - 0,49 \cdot 1,87 = 0,0837$

b) $P(5,5 < X < 5,7) = P(X < 5,7) - P(X \leq 5,5) = F(5,7) - F(5,5)$

$F(5,5) = (5,5 - 5)^2 (3 \cdot 5,5^2 - 38 \cdot 5,5 + 121) = 0,25 \cdot 2,75 = 0,6875$

$P(5,5 < X < 5,7) = 0,9163 - 0,6875 = 0,2288$

7.11

Vi bestämmer först väntevärdet och variansen för X

Antal dagar	$f(x)$	$xf(x)$	$x^2f(x)$
10	0,5	5,0	50,0
11	0,3	3,3	36,3
12	0,2	2,4	28,8
Summa	1,0	10,7	115,1

Det gäller alltså att $E(X) = 10,7$ och $V(X) = 115,1 - 10,7^2 = 0,61$

Låt $Y = 20000 + 7200X$ vara totala kostnaden för arbetet. Vi får då den förväntade totala kostnaden $E(Y) = E(20000 + 7200X) = 200000 + 7200E(X) = 277040$, variansen för den totala kostnaden $V(Y) = V(200000 + 7200X) = 7200^2V(X) = 31622400$. Standardavvikelsen för den totala kostnaden blir $\sqrt{31622400} \approx 5623,38$

7.12

Låt X vara antal tillgängliga bilar.

a) Frekvensfunktionen för X blir $f(0) = P(X = 0) = (1 - 0,9)(1 - 0,8) = 0,02$, $f(1) = P(X = 1) = 0,9(1 - 0,8) + (1 - 0,9)0,8 = 0,26$ och $f(2) = P(X = 2) = 0,9 \cdot 0,8 = 0,72$

b)

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	0,02	0,00	0,00
1	0,26	0,26	0,26
2	0,72	1,44	2,88
Summa	1,00	1,70	3,14

$E(X) = 1,70$, $V(X) = 3,14 - 1,70^2 = 0,25$ så att standardavvikelsen är $\sqrt{0,25} = 0,5$

7.13

$F(x) = 1 - e^{-x/2000}$, $x > 0$

a) $P(X > 2000) = 1 - P(X \leq 2000) = 1 - F(2000) = 1 - \left(1 - e^{-2000/2000}\right) = e^{-1} \approx 0,3679$

b) $P(1000 < X < 2000) = F(2000) - F(1000) = \left(1 - e^{-2000/2000}\right) - \left(1 - e^{-1000/2000}\right) = e^{-0,5} - e^{-1} \approx 0,2387$

7.14

a) Låt X vara lottvinsten i kronor. Vi får

x	$f(x)$	$xf(x)$	$x^2 f(x)$
0	0,7	0	0
25	0,2	5	125
50	0,06	3	150
100	0,025	2,5	250
500	0,015	7,5	3750
Summa	1,000	18	4275

$E(X) = 18$

$V(X) = 4275 - 18^2 = 3951$

b) Låt Y vara totala vinsten från fem lotter. Under förutsättning att vinsterna är stokastiskt oberoende av varandra får v i

$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{5} E(X_i) = 5E(X) = 5 \cdot 18 = 90$

$V(Y) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{5} V(X_i) = 5V(X) = 5 \cdot 3951 = 19755$

7.15

a)

x	$f(x)$	$F(x)$
0	0,30	0,30
1	0,36	0,66
2	0,22	0,88
3	0,08	0,96
4	0,03	0,99
5	0,01	1,00

b) $P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - 0,88 = 0,12$

c) Givet att en kund köper mer än två varor:

x	$f(x \mid x>2)$	$xf(x \mid x>2)$
3	0,08/0,12=8/12	24/12
4	0,03/0,12=3/12	12/12
5	0,01/0,12=1/12	5/12
Summa	1	41/12

$E(X \mid X>2) = \frac{41}{12}$

7.16

Låt X vara efterfrågan av vara A och Y vara efterfrågan av vara B. X och Y är stokastiskt oberoende.

a) $P(X=0, Y=0) = f_X(0) f_Y(0) = 0,05 \cdot 0,10 = 0,005$

b) $P(X \geq 3, Y \leq 4) = P(X \geq 3) P(X \leq 4)$

$= (f_X(3) + f_X(4) + f_X(5) + f_X(6))(f_Y(0) + f_Y(1) + f_Y(2) + f_Y(3) + f_Y(4))$

$= (0,25+0,30+0,10+0,03)(0,10+0,12+0,15+0,20+0,25) = 0,68 \cdot 0,82 = 0,5576$

c) $P(X \leq 3, Y \geq 3) = P(X \leq 3) P(X \geq 3)$

$= (f_X(0) + f_X(1) + f_X(2) + f_X(3))(f_Y(3) + f_Y(4) + f_Y(5) + f_Y(6))$

$= (0,05+0,07+0,20+0,25)(0,20+0,25+0,10+0,08) = 0,57 \cdot 0,63 = 0,3591$

d) $P((X=3, Y=0) \text{ eller } (X=0, Y=3)) = P(X=3, Y=0) + P(X=0, Y=3)$

$= 0,25 \cdot 0,10 + 0,05 \cdot 0,20 = 0,035$

e)

x	$f_X(x)$	$xf_X(x)$	$x^2 f_X(x)$	y	$f_Y(y)$	$yf_Y(y)$	$y^2 f_Y(y)$
0	0,05	0	0	0	0,10	0	0
1	0,07	0,07	0,07	1	0,12	0,12	0,12
2	0,20	0,40	0,80	2	0,15	0,30	0,60
3	0,25	0,75	2,25	3	0,20	0,60	1,80
4	0,30	1,20	4,80	4	0,25	1,00	4,00
5	0,10	0,50	2,50	5	0,10	0,50	2,50
6	0,03	0,18	1,08	6	0,08	0,48	2,88
Summa	1,00	3,10	11,50	Summa	1,00	3,00	11,90

$E(X) = 3,10$

$V(X) = 11,50 - 3,10^2 = 1,89$

$E(Y) = 3,00$

$V(Y) = 11,90 - 3,00^2 = 2,90$

f) Låt Q vara vinsten för vara A och R vara vnsten för vara B. Då är

$Q = 100X - 50(6 - X) = 150X - 300$

så att

$E(Q) = E(150X - 300) = 150E(X) - 300 = 150 \cdot 3,10 - 300 = 165$

Vidare är

$R = 100Y - 50(10 - Y) = 150Y - 500$

så att

$E(R) = E(150X - 500) = 150E(Y) - 500 = 150 \cdot 3,00 - 500 = -50$

dvs. vara A är mer lönsam.

7.17

a)

		X		
		0	1	Marginal
Y	0	0,30	0,20	0,50
	1	0,20	0,30	0,50
Marginal		0,50	0,50	1,00

Marginalfördelning för X (och för Y som är identisk med den för X)

x	$f(x)$	$xf(x)$	$x^2 f(x)$
0	0,50	0	0
1	0,50	0,50	0,50
Summa	1,00	0,50	0,50

b) Om $X = 0$ blir betingade fördelningen för Y:

y	$f(y \mid x=0)$	$yf(y \mid x=0)$	$y^2 f(y \mid x=0)$
0	0,30/0,50=0,60	0	0
1	0,20/0,50=0,40	0,40	0,40
Summa	1,00	0,40	0,40

Om $X = 1$ blir betingade fördelningen för Y:

y	$f(y \mid x=1)$	$yf(y \mid x=1)$	$y^2 f(y \mid x=1)$
0	0,20/0,50=0,40	0	0
1	0,30/0,50=0,60	0,60	0,60
Summa	1,00	0,60	0,60

c) $E(Y \mid X=0) = 0,4$ och $E(Y \mid X=1) = 0,6$ (enligt tabeller i uppgift b)

d) Vi kan sluta oss till att X och Y är stokastiskt beroende eftersom t.ex. $E(Y \mid X=0) \neq E(Y \mid X=1)$.

e) $xyf(x,y)$:

		X	
		0	1
Y	0	0,00	0,00
	1	0,00	0,30

så att $E(XY) = 0,30$

I a) ser vi att $E(X) = E(Y) = 0,50$ och $V(X) = V(Y) = E(X^2) - (E(X))^2 = 0,50 - 0,50^2 = 0,25$ så att

$Cov(X,Y) = E(XY) - E(X)E(Y) = 0,30 - 0,5 \cdot 0,5 = 0,05$

$\rho = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{0{,}05}{\sqrt{0{,}25 \cdot 0{,}25}} = 0,2$

f)

z	$f(z)$	$zf(z)$	$z^2 f(z)$
0	0,30	0	0
1	0,20+0,20=0,40	0,40	0,40
2	0,30	0,60	1,20
Summa	1,00	1,00	1,60

g) Enligt tabell i uppgift f får vi $E(Z) = 1,00$ och $V(Z) = E(Z^2) - (E(Z))^2 = 1,60 - 1,00^2 = 0,60$

Med hjälp av räkneregler för väntevärden och varianser för linjära kombinationer får vi

$E(Z) = E(X+Y) = E(X) + E(Y) = 0,50 + 0,50 = 1,00$

$V(Z) = V(X+Y) = V(X)+V(Y)+2Cov(X,Y) = 0,25+0,25+2\cdot 0,05 = 0,60$

7.18

Låt X vara antal sålda partier muttrar och Y vara antal sålda partier skruvar.

a) Vi söker $P(X+Y=3)$

$= P(X=3 \cap Y=0)+P(X=2 \cap Y=1)+P(X=1 \cap Y=2)+P(X=0 \cap Y=3)$

$= 0,00 + 0,04 + 0,10 + 0,00 = 0,14$

b) Låt T vara vinsten av försäljningen av mutrar och U vara vinsten av försäljningen av skruvar. Fördelningen av T och U är

			T		
		0	3000	6000	9000
	0	0,05	0,10	0,05	0,00
U	5000	0,05	0,15	0,04	0,01
	10000	0,00	0,10	0,20	0,05
	15000	0,00	0,05	0,05	0,10

Vi söker $P(T > U)$

$= P(X=1, Y=0)+P(X=2, Y=0)+P(X=2, Y=1)+P(X=3, Y=0)+P(X=3, Y=1)$

$= 0,10 + 0,05 + 0,04 + 0,00 + 0,01 = 0,20$

c) $E(T+U) = E(T) + E(U)$

$= 0,10 \cdot 0 + 0,40 \cdot 3000 + 0,34 \cdot 6000 + 0,16 \cdot 9000 + 0,20 \cdot 0 + 0,25 \cdot 5000 + 0,35 \cdot 10000 + 0,20 \cdot 15000$

$= 1200 + 2040 + 1440 + 1250 + 3500 + 3000 = 12430$

d)

x	$f(x)$	$xf(x)$	y	$f(y)$	$yf(y)$
0	0,10	0,00	0	0,20	0,00
1	0,40	0,40	1	0,25	0,25
2	0,34	0,68	2	0,35	0,70
3	0,16	0,48	3	0,20	0,60
Summa	1,00	1,56	Summa	1,00	1.55

e) $E(X) = 1,56$, $E(Y) = 1,55$

f)

x	$f(x \mid y=0)$	$f(x \mid y=1)$	$f(x \mid y=2)$	$f(x \mid y=3)$
0	0,05/0,20=0,25	0,05/0,25=0,20	0,00/0,35=0,00	0,00/0,20=0,00
1	0,10/0,20=0,50	0,15/0,25=0,60	0,10/0,35≈0,29	0,05/0,20=0,25
2	0,05/0,20=0,25	0,04/0,25=0,16	0,20/0,35≈0,57	0,05/0,20=0,25
3	0,00/0,20=0,00	0,01/0,25=0,04	0,05/0,35≈0,14	0,10/0,20=0,50
Summa	1,00	1,00	1,00	1,00

g)

x	$f(x \mid y=0)$	$xf(x \mid y=0)$	x	$f(x \mid y=1)$	$xf(x \mid y=1)$
0	0,25	0,00	0	0,20	0,00
1	0,50	0,50	1	0,60	0,60
2	0,25	0,50	2	0,16	0,32
3	0,00	0,00	3	0,04	0,12
Summa	1,00	1,00	Summa	1,00	1,04

x	$f(x \mid y=2)$	$xf(x \mid y=2)$	x	$f(x \mid y=3)$	$xf(x \mid y=3)$
0	0,00	0,00	0	0,00	0,00
1	0,29	0,29	1	0,25	0,25
2	0,57	1,14	2	0,25	0,50
3	0,14	0,42	3	0,50	1,50
Summa	1,00	1,85	Summa	1,00	2,25

$E(X \mid Y=0) = 1,00$, $E(X \mid Y=1) = 1,04$, $E(X \mid Y=2) = 1,85$, $E(X \mid Y=3) = 2,25$

h) Försäljningen av skruvar och muttrar är beroende av varandra eftersom, t.ex. det betingade väntevärdet för anal sålda partier muttrar givet antal sålda partier skruvar är stokastiskt beroende av antal sålda partier skruvar.

7.19

a) Låt X vara vinsten vid investering A och Y vara vinsten vid investering B. Då är simultana fördelningen$X$ och Y

		X		
		-1000	2000	Marginal
Y	-1000	0,15	0,25	0,40
	2000	0,45	0,15	0,60
Marginal		0,60	0,40	1,00

b)

Väntevärden och varianser för X och Y bestäms genom

x	$f(x)$	$xf(x)$	$x^2f(x)$	y	$f(y)$	$yf(y)$	$y^2f(y)$
-1000	0,60	-600	600000	-1000	0,40	-400	400000
2000	0,40	800	1600000	2000	0,60	1200	2400000
Summa	1,00	200	2200000	Summa	1,00	800	2800000

$E(X) = 200$, $V(X) = E(X^2) - (E(X))^2 = 2200000 - 200^2 = 2160000$

$E(Y) = 800$, $V(Y) = E(Y^2) - (E(Y))^2 = 2800000 - 800^2 = 2160000$

I följande tabell visas $xyf(x, y)$:

		X	
		-1000	2000
Y	-1000	150000	-500000
	2000	-900000	600000

$E(XY) = \sum xyf(x, y) = -650000$

$Cov(X, Y) = E(XY) - E(X)E(Y) = -65000 - 200 \cdot 800 = -810000$

$\rho = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{-810000}{\sqrt{2160000 \cdot 2160000}} = -0,375$

c) $E(X + Y) = E(X) + E(Y) = 200 + 800 = 1000$

$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y) = 2160000 + 2160000 - 2 \cdot 810000 = 2700000$

7.20

a)

		Basenhet			
		HC	VHC	UHC	
Hård-	60 GB	0,2	0,1	0,1	0,4
disk	124 GB	0,1	0,2	0,3	0,6
		0,3	0,3	0,4	1,0

$P(HC) \cdot P(60GB) = 0,3 \cdot 0,4 = 0,12 \neq 0,2 = P(HC \cap 60GB)$ dvs. valet av basenhet och hårddisk är stokastiskt beroende av varandra.

b)

Kombination	Pris	Sannolikhet
HC + 60 GB	7000	0,2
HC + 124 GB	8000	0,1
VHC + 60 GB	8000	0,1
VHC + 124 GB	9000	0,2
UHC + 60 GB	9000	0,1
UHC + 124 GB	10000	0,3

Låt X vara priset för ett datorpaket. Sannolikhetsfördelningen för X blir då

x	$f(x)$	$xf(x)$	$x^2 f(x)$
7000	0,2	1400	9800000
8000	0,2	1600	12800000
9000	0,3	2700	24300000
10000	0,3	3000	30000000
Summa	1,0	8700	76900000

d) $E(X) = 8700$, $V(X) = 76900000 - 8700^2 = 12100000$

7.21

a)

		USB minnen			
		0	1	2	
Kabel	0	0,12	0,12	0,16	0,40
	1	0,18	0,18	0,24	0,60
		0,30	0,30	0,40	1,00

b) Låt X vara prset på köpta varor

x	$f(x)$	$F(x)$	$xf(x)$	$x^2 f(x)$
0	0,12	0,12	0	0
50	0,30	0,42	15	750
100	0,34	0,76	34	3400
150	0,24	1,00	36	5400
Summa	1,00		85	9550

d) $E(X) = 85$

$V(X) = E(X^2) - (E(X))^2 = 9550 - 85^2 = 2325$

7.22

a) Låt X vara priset på gammaaktien och Y vara priset på deltaaktien.

x	$f(x)$	$xf(x)$	$x^2 f(x)$	y	$f(y)$	$yf(y)$	$y^2 f(y)$
40	0,2	8,0	320,0	41	0,1	4,1	168,1
42	0,7	29,4	1234,8	43	0,5	21,5	924,5
45	0,1	4,5	202,5	44	0,4	17,6	774,4
Summa	1,0	41,9	1757,3	Summa	1,0	43,2	1867,0

$E(X) = 41,9,\ V(X) = 1757,3 - 41,9^2 = 1,69$

$E(Y) = 43,2,\ V(Y) = 1867,0 - 43,2^2 = 0,76$

b) Eftersom X och Y är stokastiskt oberoende blir simultana sannolikheterna lika med produkten av marginalsannolikheterna.

			x	
		40	42	45
	41	0,02	0,07	0,01
y	43	0,10	0,35	0,05
	44	0,08	0,28	0,04

c) $P(X > Y) = P(X = 42, Y = 41) + P(X = 45, Y = 41) + P(X = 45, Y = 43) + P(X = 45, Y = 44)$

$= 0,07 + 0,01 + 0,05 + 0,04 = 0,17$

d) $E(5X + 8Y) = 5E(X) + 8E(Y) = 5 \cdot 41,9 + 8 \cdot 43,2 = 209,5 + 345,6 = 555,1$

$V(5X + 8Y) = 5^2 V(X) + 8^2 V(Y) = 25 \cdot 1,69 + 64 \cdot 0,76 = 90,89$

7.23

a) Låt X vara priset på gammaaktien och Y vara priset på deltaaktien

x	$f(x)$	$xf(x)$	$x^2 f(x)$	y	$f(y)$	$yf(y)$	$y^2 f(y)$
40	0,20	8,00	320	41	0,10	4,10	168,1
42	0,70	29,40	1234,8	43	0,50	21,50	924,5
45	0,10	4,50	202,5	44	0,40	17,60	774,4
Summa	1,00	41,90	1757,3	Summa	1,00	43,20	1867

$E(X) = 41,90,\ V(X) = E(X^2) - (E(X))^2 = 1757,3 - 41,90^2 = 1,69$

$E(Y) = 43,20,\ V(Y) = E(Y^2) - (E(Y))^2 = 1867 - 43,20^2 = 0,76$

b) $P(X = 42)\,P(Y = 43) = 0,70 \cdot 0,50 = 0,35$ men $P(X = 42 \cap Y = 43) = 0,40$ så att $P(X = 42)\,P(Y = 43) \neq P(X = 42 \cap Y = 43)$ och det följer att priserna är stokastiskt beroende.

c) $P(X > Y) = 0,02 + 0,06 + 0,04 + 0,00 = 0,12$

d) $xyf(x,y)$:

	40	42	45
41	38,20	34,44	110,70
43	103,20	722,40	77,40
44	211,20	517,44	0,00

$E(XY) = 1809,58$

$Cov(X,Y) = E(XY) - E(X)E(Y) = 1809,58 - 41,90 \cdot 43,20 = -0,5$

$E(5X+8Y) = 5E(X) + 8E(Y) = 5 \cdot 41,90 + 8 \cdot 43,20 = 555,10$

$V(5X+8Y) = 5^2V(X) + 8^2V(Y) + 2 \cdot 5 \cdot 8Cov(X,Y) = 25 \cdot 1,69 + 64 \cdot 0,76 - 80 \cdot 0,5 = 50,89$

7.24

a)

		AJ			
		-50	0	100	Marg.
	-100	0,14	0,06	0,10	0,30
CD	0	0,02	0,03	0,05	0,10
	100	0,08	0,12	0,10	0,30
	150	0,06	0,09	0,15	0,30
	Marg.	0,40	0,30	0,40	1,00

b) $P(AJ = -50)P(CD = -100) = 0,40 \cdot 0,30 = 0,12$

$\neq 0,14 = P(AJ = -50 \cap CD = -100)$, dvs. AJ och CD är stokastiskt beroende.

c)

CD	$f(cd)$	$cdf(cd)$	$cd^2f(cd)$
-100	0,30	-30	3000
0	0,10	0	0
100	0,30	30	3000
150	0,30	45	6750
Summa	1,00	45	12750

$E(CD) = 45$, $V(CD) = 12750 - 45^2 = 10725$

d) De betingade fördelningarna för CD givet olika värden på AJ är olika för olika värden på AJ eftersom CD och AJ är stokastiskt beroende.

7.25

a) Låt X beteckna pris och Y försäljningsvolym

x	$f(x)$
30	0,30
35	0,40
40	0,30
Summa	1,00

y	$f(y)$
30	0,15
40	0,25
50	0,30
60	0,30
Summa	1,00

b)

y	$f(y \mid x=30)$	$yf(y \mid x=30)$
30	0,00/0,30=0	0
40	0,05/0,30=1/6	40/6
50	0,10/0,30=1/3	50/3
60	0,15/0,30=1/2	60/2
Summa	1,00	320/6$\approx 53,33$

y	$f(y \mid x=35)$	$yf(y \mid x=35)$
30	0,05/0,40=1/8	30/8
40	0,10/0,40=1/4	40/4
50	0,15/0,40=3/8	150/8
60	0,10/0,40=1/4	60/4
Summa	1,00	380/8=47,50

y	$f(y \mid x=40)$	$yf(y \mid x=40)$
30	0,10/0,30=1/3	30/3
40	0,10/0,30=1/3	40/3
50	0,05/0,30=1/6	50/6
60	0,05/0,30=1/6	60/6
Summa	1,00	250/6$\approx 41,67$

c) Se tabell i uppgift b. $E(Y \mid X=30) \approx 53,33$, $E(Y \mid X=35) = 47,50$ och $E(Y \mid X=40) \approx 41,67$

d) Det betingade väntevärdet försäljningsvolym är olika för olika priser, alltså är försäljningsvolym och pris stokastiskt beroende.

7.26

b, c)

		X			
		0	1	2	
	0	1/9	2/9	1/9	4/9
Y	1	2/9	2/9	0	4/9
	2	1/9	0	0	1/9
		4/9	4/9	1/9	1

7.27

a)

				y			
		0	1	2	3	4	Marg.
	1	0,08	0,13	0,09	0,06	0,03	0,39
x	2	0,03	0,08	0,08	0,09	0,07	0,35
	3	0,01	0,03	0,06	0,08	0,08	0,26
	Marg.	0,12	0,24	0,23	0,23	0,18	1,00

b)

y	$f(y \mid x=3)$	$yf(y \mid x=3)$
0	0,01/0,26=$\frac{1}{26}$	0
1	0,03/0,26=$\frac{3}{26}$	3/26
2	0,06/0,26=$\frac{6}{26}$	12/26
3	0,08/0,26=$\frac{8}{26}$	24/26
4	0,08/0,26=$\frac{8}{26}$	32/26
Summa	1,00	71/26

c) $E(Y \mid X = 3) = 71/26$

d) $P(X = 1) \cdot P(Y = 0) = 0,39 \cdot 0,12 = 0,0468 \neq 0,08 = P(X = 1 \cap Y = 0)$, dvs. X och Y är stokastiskt beroende.

8. NÅGRA VANLIGA DISKRETA SANNOLIKHETSGFÖRDELNINGAR

8.1

a) 0,2048; b) 0,2048; c) 0,2150; d) 0,0739

8.2

Låt X vara antal lampor som lyser efter 1000 timmar och Y vara antal lampor som inte lyser efter 1000 timmar. Då är $X \sim bin\,(n=3, p=0,9)$ och $Y \sim bin\,(n=3, p=0,1)$.

a) $P\,(X=0) = P\,(Y=3) = 0,0010$ b) $P\,(X=1) = P\,(Y=2) = 0,0270$ c) $P\,(X \geq 2) = P\,(Y \leq 1) = P\,(Y=0) + P\,(Y=1) = 0,7290 + 0,2430 = 0,9720$

8.3

Låt X vara antal personer som väljer 3 eller 7. Om alla siffror väljs slumpmässigt är då $X \sim bin\,(n=10, p=0,2)$. Vi söker $P\,(X \geq 4) = 1 - P\,(X \leq 3) = 1 - 0,8791 = 0,1209$

8.4

Om det inte är någon skillnad melan produkterna är sannolikheten 0,5 att ett hushåll föredrar produkt A. Låt X vara antal hushåll som föredrar produkt A. Då är $X \sim bin\,(n=15, p=0,5)$.

a) Vi söker $P\,(X \geq 10) = 1 - P\,(X \leq 9) \approx 1 - 0,849 = 0,151$

b) $P\,(X \leq 5 \text{ eller } X \geq 10) = P\,(X \leq 5) + P\,(X \geq 10) = 0,151 + 0,151 = 0,302$

8.5

Låt X vara antal personer i urvalet som har minst en tv-apparat. Då är $X \sim bin\,(n=12, p=0,75)$. Vi söker $P\,(X=10) = 0,2323$

8.6

Låt X vara antal felaktiga enheter som kan epareras. Då är $X \sim bin\left(n=6, p=0,40\right)$. Vi söker $P\left(X \geq 3\right) = 1 - P\left(X \leq 2\right) = 1 - 0,544 = 0,456$

8.7

Låt X vara antalet felaktiga enheter i ett urval. Då är $X \sim bin\left(n=10, p\right)$. Ett parti accepteras om $x \leq 2$.

a) $p = 0,02$. $P\left(\text{partiet accepteras}\right) = P\left(X \leq 2\right) = 0,9991$

b) $p = 0,05$. $P\left(\text{partiet accepteras}\right) = P\left(X \leq 2\right) = 0,9885$

c) $p = 0,10$. $P\left(\text{partiet accepteras}\right) = P\left(X \leq 2\right) = 0,9298$

d) Nu är $X \sim bin\left(n=20, p=0.10\right)$ och ett parti accepteras om $x \leq 4$

Då blir $P\left(\text{partiet accepteras}\right) = P\left(X \leq 4\right) = 0,9568$

8.9

Låt X vara antal snurrningar man måste göra med hjulet tills man får rött för första gången. Då är X geometriskt fördelad med $p = \frac{18}{38} = \frac{9}{19}$. Frekvensfunktionen är då $f\left(x\right) = \left(\frac{10}{19}\right)^{x-1}\frac{9}{19}$, för $x = 1, 2, \ldots$ $E\left(X\right) = \frac{1}{p} = \frac{19}{9}$. $P\left(X > \frac{19}{9}\right) = P\left(X \geq 3\right) = 1 - P\left(X \leq 2\right) = 1 - \left(P\left(X=1\right) + P\left(X=2\right)\right) = 1 - \left(\left(\frac{10}{19}\right)^{1-1}\frac{9}{19} + \left(\frac{10}{19}\right)^{2-1}\frac{9}{19}\right) = 1 - \left(\frac{9}{19} + \frac{90}{361}\right) = 1 - \frac{171+90}{361} = \frac{100}{361}$

8.10

Låt X_1 vara antal personer i urvalet i åldern 18-24 år, X_2 antal personer i åldern 25-34 år, X_3 antal personer i åldern 35-44 år, X_4 antal personer i åldern 45-64 år och X_5 antal personer över 64 år. Då är $(X_1, X_2, X_3, X_4, X_5)$ multinomialfördelad med parametrarna $n = 5, p_1 = 0,21, p_2 = 0,24, p_3 = 0,14, p_4 = 0,26, p_5 = 0,15$. Vi söker

$P\left(X_1 = 0, X_2 = 2, X_3 = 0, X_4 = 2, X_5 = 1\right) = \frac{5!}{0!2!0!2!1!}0,21^0 0,24^2 0,14^0 0,26^2 0,15^1 = 30 \cdot 0,0576 \cdot 0,0676 \cdot 0,15 = 0,01752$

8.11

Låt X_1 vara antal enheter i urvalet med exakt ett fel, X_2 antal enheter med mer än ett fel och X_3 antal felfria enheter.

a) $X_2 \sim bin\,(n = 10, p_1 = 0,05)$. Vi söket $P\,(X_2 \geq 1) = 1 - P\,(X_2 = 0) = 1 - 0,95^{10} \approx 0,4013$

b) (X_1, X_2, X_3) är multinomialfördelad med $n = 10$, $p_1 = 0,10$, $p = 0,05$ och $p_3 = 0,85$. Vi söker $P\,(X_1 = 1, X_2 = 1, X_3 = 8) = \frac{10!}{1!1!8!} 0,10^1 0,05^1 0,85^8 \approx 0,1226$

c) $E\,(X_1) = np_1 = 10 \cdot 0,10 = 1$, $V\,(X_1) = np_1\,(1 - p_1) = 10 \cdot 0,10 \cdot 0,9 = 0,9$

$E\,(X_2) = np_2 = 10 \cdot 0,05 = 0,5$, $V\,(X_2) = np_2\,(1 - p_2) = 10 \cdot 0,05 \cdot 0,95 = 0,475$

$Cov\,(X_1, X_2) = -np_1p_2 = -10 \cdot 0,1 \cdot 0,05 = -0,05$

$E\,(2X_1 + 5X_2) = 2E\,(X_1) + 5E\,(X_2) = 2 \cdot 1 + 5 \cdot 0,5 = 4,5$

$V\,(2X_1 + 5X_2) = 2^2V\,(X_1) + 5^2V\,(X_2) + 2 \cdot 2 \cdot 5Cov\,(X_1, X_2) = 4 \cdot 0,9 + 25 \cdot 0,475 - 20 \cdot 0,05 = 14,475$

8.12

Låt X vara antal besök på hemsidan. Vi vet att $X \sim Po\,(\lambda = 4)$.

a) Vi söker $P\,(X = 0) = 0,0183$

b) Vi söker $P\,(X > 3) = 1 - P\,(X \leq 3) = 1 - 0,4335 = 0,5665$

8.13

Låt X vara antal gånger maskinen går sönder under en månad. Vi vet att $X \sim Po\,(\lambda = 3)$. Vi söker $P\,(X \geq 6) = 1 - P\,(X \leq 5) \approx 1 - 0,9161 = 0,0839$ vilket inte kan anses vara ovanligt.

8.14

Låt X vara antal träffar för Sol och Y vara antal träffar för Viller. Då är $X \sim bin\,(n = 4, p = 0,5)$ och $Y \sim bin\,(n = 4, p = 0,4)$. Eftersom X och Y är stokastiskt oberoende blir deras simultana sannolikhetsfördelning

		x						
		0	1	2	3	4	Summa
	0						0,1296
	1	0,0216					0,3456
y	2	0,0216	0,0864				0,3456
	3	0,0096	0,0384	0,0576			0,1536
	4	0,0016	0,0064	0,0096	0,0064		0,0256
	Summa	0,0625	0,2500	0,3750	0,2500	0,0625	1.0000

Vi söker $P(X < Y) = 0.0216 + 0,0216 + 0,0864 + 0,0096 + 0,0384 + 0,0576 + 0,0016 + 0,0064 + 0,0096 + 0,0064 = 0,2592$

8.15

Låt X vara antal apor i experimentet som utvecklar immunförsvaret. Då är $X \sim bin(n = 18, p)$. Utvecklingsarbetet fortsätter om $x \geq 10$.

a) $P(\text{utvecklingsarbetet läggs ner}; p = 0,6) = P(X < 10) = P(X \leq 9) \approx 0,2632$

b) $P(\text{utvecklingsarbetet fortsätter}; p = 0,5) = P(X \geq 10) = 1 - P(X \leq 9) \approx 0,4073$

8.16

Låt X vara antal fungerande tillverkningslinjer och Y vara antal tillverkningslinjer som är ur funktion. Då är $X \sim bin(n = 10, p = 0,9)$ och $Y \sim bin(n = 10, p = 0,1)$.

a) Vi söker $P(X = 10) = P(Y = 0) = 0,3487$

b) Vi söker $P(Y \leq 2) = 0,9298$

8.17

Låt X vara antal småhus med garage i urvalet. Vi vet att $X \sim bin(n = 10, p = 0,40)$

a) $P(X = 2) \approx 0,1209$

b) $P(X \leq 2) \approx 0,1673$

8.18

Låt X vara antal dagar som kursen på ST-aktien går ner eller är oförändrad bland 5 slumpmässigt valda dagar. Då är $X \sim bin\,(n = 5, p = 0,60)$.

a) Vi söker $P\,(X = 2) = 0,2304$

b) Vi söker $P\,(X \leq 2) = 0,3174$

8.19

Låt X vara antal felaktiga enheter i urvalet. Vi vet att $X \sim bin\,(n = 50, p)$. Partiet accepteras om $x \leq 2$

a) $P\,(\text{partiet accepteras; } p = 0,03) = P\,(X \leq 2) \approx 0,8108$

b) $P\,(\text{partiet accepteras; } p = 0,10) = P\,(X \leq 2) \approx 0,1117$

8.20

a) Vi vet att $f\,(0) = 0,605$, dvs. $\frac{\lambda^0 e^{-\lambda}}{0!} = 0,605$. Detta innebär att $e^{-\lambda} = 0,605$, $-\lambda = \ln 0,605$ och $\lambda = 0,5025$

b) $f\,(1) = \frac{\lambda^1 e^{-\lambda}}{1!} = 0,5025 \cdot 0,605 = 0,3040$

c) Låt X vara antal dagar med exakt ett ärende. Då är $X \sim bin\,(n = 10, p = 0,3040)$. Vi söker $P\,(X = 3) = 0,2667$.

9. NÅGRA VANLIGA KONTINUERLIGA SANNOLIKHETSGFÖRDELNINGAR

9.3

Vi vet att $X \sim N\left(\mu = 50, \sigma^2 = 100\right)$.

a) $P\left(X \leq 65\right) = P\left(\frac{X-\mu}{\sigma} \leq \frac{65-50}{10}\right) = P\left(Z \leq 1,5\right) = \Phi\left(1,5\right) = 0,9332.$

b) $P\left(X \leq 25\right) = P\left(\frac{X-\mu}{\sigma} \leq \frac{25-50}{10}\right) = P\left(Z \leq -2,5\right) = \Phi\left(-2,5\right) = 0,0062.$

c) $P\left(X > 35\right) = P\left(\frac{X-\mu}{\sigma} > \frac{35-50}{10}\right) = P\left(Z > -1,5\right) = 1 - \Phi\left(-1,5\right) = 0,9332.$

d) $P\left(X > 70\right) = P\left(\frac{X-\mu}{\sigma} > \frac{70-50}{10}\right) = P\left(Z > 2\right) = 1 - \Phi\left(2\right) = 0,0228.$

e) $P\left(40 < X < 60\right) = P\left(\frac{40-50}{10} < \frac{X-\mu}{\sigma} < \frac{60-50}{10}\right) = P\left(-1 < Z < 1\right)$

$= \Phi\left(1\right) - \Phi\left(-1\right) = 0,6827.$

f) $P\left(|X-50| \leq 20\right) = P\left(30 \leq X \leq 70\right) = P\left(\frac{30-50}{10} < \frac{X-\mu}{\sigma} < \frac{70-50}{10}\right)$

$= P\left(-2 < Z < 2\right) = \Phi\left(2\right) - \Phi\left(-2\right) = 0,9544.$

g) $P\left(|X-50| > 15\right) = P\left(X < 35 \text{ eller } X > 65\right) = P\left(X < 35\right) + P\left(X > 65\right) = P\left(\frac{X-\mu}{\sigma} < \frac{35-50}{10}\right) + P\left(\frac{X-\mu}{\sigma} > \frac{65-50}{10}\right) = P\left(Z < 1,5\right) + P\left(Z < -1,5\right) = \Phi\left(1,5\right) + \left(1 - \Phi\left(1,5\right)\right) = 0,1336.$

9.4

a) $z > 0,67$ b) $z > -0,67$ c) $z < -0,67$ d) $-1,64 < z < 1,64$

9.5

Låt X vara taxeringsvärdet på ett småhus i området. Vi vet att $X \sim N\left(\mu = 2500000, \sigma^2 = 300000^2\right)$. Vi söker $P\left(X > 2800000\right) = P\left(\frac{X-\mu}{\sigma} > \frac{2800000-2500000}{300000}\right)$ $P\left(Z > 1\right) = 1 - \Phi\left(1\right) = 0,1587$

9.6

Låt X vara kolesterolvärdet hos en slumpmässigt vald kvinna. Enligt modellen är då $X \sim N\left(\mu = 6,5, \sigma^2 = 0,75^2\right)$.

a) $P(X > 7) = 1 - P(X \leq 7) = 1 - P\left(\frac{X-6,5}{0,75} \leq \frac{7-6,5}{0,75}\right) = 1 - P(Z \leq 0,67) = 1 - \Phi(0,67) \approx 1 - 0,7486 = 0,2514$

b) $P(5 \leq X \leq 6) = P(X \leq 6) - P(X \leq 5) = P\left(\frac{X-6,5}{0,75} \leq \frac{6-6,5}{0,75}\right) - P\left(\frac{X-6,5}{0,75} \leq \frac{5-6,5}{0,75}\right) = P(Z \leq -0,67) - P(Z \leq -2) \approx 0,2514 - 0,0228 = 0,2286$

9.7

Låt X vara vikten i gram hos en slumpmässigt vald konservburk. Det gäller att $X \sim N\left(\mu, \sigma^2 = 10^2\right)$. Vi söker μ sådant att $P(X \leq 300) \leq 0,05$. Men då är $P\left(\frac{X-\mu}{\sigma} \leq \frac{300-\mu}{10}\right) \leq 0,05$, dvs. $\Phi\left(\frac{300-\mu}{10}\right) \leq 0,05$. Eftersom $\Phi(-1,64) = 0,05$ följer det att $\frac{300-\mu}{10} \leq -1,64$. Lösningen till denna ekvation är $\mu \geq 316,4$, dvs. det minsta värdet på μ är 316,4.

9.8

a) Låt X_1 vara resultatet från projekt 1 och vi låter $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$. Vi vet att $P(X_1 > 100000) = 0,1$ dvs. $P\left(\frac{X_1-\mu_1}{\sigma_1} > \frac{100000-\mu_1}{\sigma_1}\right) = 0,1$ och $P\left(Z > \frac{100000-\mu_1}{\sigma_1}\right) = 0,1$. Eftersom $P(Z > 1,28) = 0,1$ ser vi att $\frac{100000-\mu_1}{\sigma_1} = 1,28$.Vi vet också att $P(X_1 < 0) = 0,1$ dvs. $P\left(\frac{X_1-\mu_1}{\sigma_1} < \frac{0-\mu_1}{\sigma_1}\right) = 0,1$ och $P\left(Z < -\frac{\mu_1}{\sigma_1}\right) = 0,1$. Eftersom $P(Z < -1,28) = 0,1$ ser vi att $-\frac{\mu_1}{\sigma_1} = -1,28$. Vi har därmed två ekvationer och två obekanta. Lösningen blir $\mu_1 = 50000$, $\sigma_1 = 39062,5$.

Låt nu X_2 vara resultatet från projekt 2 och vi låter $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$. Vi vet att $P(X_2 > 150000) = 0,1$ dvs. $P\left(\frac{X_2-\mu_2}{\sigma_2} > \frac{150000-\mu_2}{\sigma_2}\right) = 0,1$ och $P\left(Z > \frac{150000-\mu_2}{\sigma_2}\right) = 0,1$. Eftersom $P(Z > 1,28) = 0,1$ ser vi att $\frac{150000-\mu_2}{\sigma_2} = 1,28$.Vi vet också att $P(X_2 < 0) = 0,1$ dvs. $P\left(\frac{X_2-\mu_2}{\sigma_2} < \frac{0-\mu_2}{\sigma_2}\right) = 0,1$ och $P\left(Z < -\frac{\mu_2}{\sigma_2}\right) = 0,1$. Eftersom $P(Z < -1,28) = 0,1$ ser vi att $-\frac{\mu_2}{\sigma_2} = -1,28$. Vi har därmed även här två ekvationer och två obekanta. Lösningen blir $\mu_2 = 75000$, $\sigma_2 = 58593,75$.

b) Förväntade vinsten för båda projekten tillsammans är $E(X_1 + X_2) =$

$E(X_1) + E(X_2) = \mu_1 + \mu_2 = 50000 + 75000 = 125000$

Variansen för vinsten från båda projekten tillsammans är $V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2) = \sigma_1^2 + \sigma_2^2 = 39062,5^2 + 58593,75^2 = 4959106445,3125$. Observera att $Cov(X_1, X_2) = 0$ på grund av oberoende, Standardavvikelsen för vinsten från båda projekten tillsammans blit därför $\sqrt{4959106445,3125} \approx 70420,92$

9.10

$X \sim bin(n = 80, p = 1/3)$. Vi ser att $np = 80 \cdot \frac{1}{3} \approx 26,67 > 5$ och $n(1-p) = 80 \cdot \frac{2}{3} \approx 53,33$ så approximation med normalfördelningen kan bli bra, där $\mu = np = 80 \cdot \frac{1}{3} \approx 26,67$ och $\sigma^2 = np(1-p) = 80 \cdot \frac{1}{3} \cdot \frac{2}{3} \approx 17,78$.

a) $P(X = 27) = P(26,5 < X < 27,5) = P\left(\frac{26,5-26,67}{\sqrt{17,78}} < \frac{X-\mu}{\sigma} < \frac{27,5-26,67}{\sqrt{17,78}}\right) \approx P(-0,04 < Z < 0,20) = \Phi(0,20) - \Phi(-0,04) = 0,5793 - 0,4840 = 0,0953$

b) $P(X \leq 20) = P(X \leq 20,5) = P\left(\frac{X-\mu}{\sigma} \leq \frac{20,5-26,67}{\sqrt{17,78}}\right) \approx P(Z \leq -1,46) = \Phi(-1,46) = 0,0722$

c) $P(25 \leq X \leq 28) = P(24,5 \leq X \leq 28,5) = P\left(\frac{24,5-26,67}{\sqrt{17,78}} \leq \frac{X-\mu}{\sigma} \leq \frac{28,5-26,67}{\sqrt{17,78}}\right) \approx P(-0,51 \leq Z \leq 0,43) = \Phi(0,43) - \Phi(-0,51) = 0,6664 - 0,3050 = 0,3614$

9.11

Låt X vara antalet personer i urvalet som känner till produkten. Om tillverkarens påstående är riktigt så är $X \sim bin(n = 1000, p = 0,20)$.

a) $E(X) = np = 1000 \cdot 0,20 = 200$, $V(X) = np(1-p) = 1000 \cdot 0,20 \cdot 0,80 = 160$ Standardavvikelsen är $\sqrt{160} \approx 12,65$

b) $P(X \leq 150) = P(X \leq 150 + 0,5) = P\left(\frac{X-E(X)}{\sqrt{V(X)}} \leq \frac{150+0,5-200}{12,65}\right) = P(Z \leq -3,91)$

$= \Phi(-3,91) \approx 0$

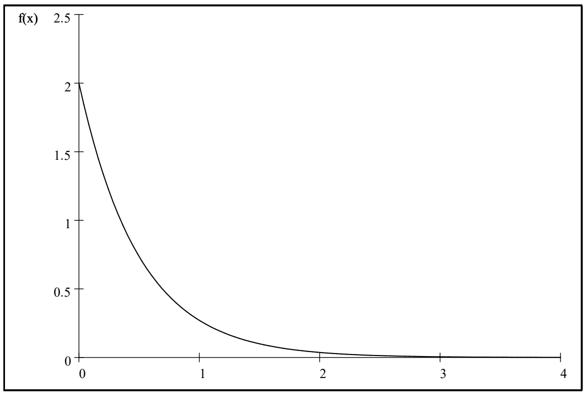
c) Om reklamfirmans påstående är sant är det enligt uppgift b) mycket osannolikt att 150 personer eller färe bland 1000 slumpmässigt valda personer känner till den nya produkten. Det är därför inte rimligt att tro på reklamfirmans påstående är sant.

9.12

$X \sim \exp(\lambda = 2)$.

a) X kan anta alla positiva reella tal.

b)



c) $P(X \leq 1) = F(1) = 1 - e^{-2 \cdot 1} \approx 0,864\,7$

d) $P(0,25 \leq X \leq 1) = F(1) - F(0,25) = (1 - e^{-2 \cdot 1}) - (1 - e^{-2 \cdot 0,25}) \approx 0,4712$

e) $P(X \geq 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - (1 - e^{-2 \cdot 2}) \approx 0,0183$

f) $E(X) = \frac{1}{\lambda} = \frac{1}{2} = 0,5$, $V(X) = \frac{1}{\lambda^2} = \frac{1}{2^2} = 0,25$ så att standardavvikelsen är $\sqrt{0,25} = 0,5$

g) $P(|X - 0,5| < 0,5) = P(0 < X < 1) = F(1) \approx 0,8647$

9.13

Låt X vara tiden i timmar mellan två driftstopp. Vi vet att antal driftstopp per 500 timmar är Poissonfördelad med väntevärdet 2, dvs. det inträffar i medeltal 2/500=0,004 driftstopp per timme. Detta innebär att $X \sim \exp(\lambda = 0,004)$.

a) $E(X) = \frac{1}{\lambda} = \frac{1}{0,004} = 250$.

b) Vi söker $P(100 < X < 300) = F(300) - F(100) = (1 - e^{-0,004 \cdot 300}) - (1 - e^{-0,004 \cdot 100}) = e^{-0,4} - e^{-1,2} \approx 0,3691$.

9.14

Låt X vara mängden blekmedel i en slumpmässigt vald flaska. Då är $X \sim N(\mu, \sigma^2 = 0,2^2)$. Vi söker μ så att $P(X > 2,46) = 0,01$. Men då är $P\left(\frac{X-\mu}{\sigma} > \frac{2,46-\mu}{0,2}\right) = 0,01$, dvs. $P\left(Z > \frac{2,46-\mu}{0,2}\right) = 0,01$. Eftersom $P(Z > 2,33) = 0,01$ ser vi att $\frac{2,46-\mu}{0,2} = 2,33$, dvs. $\mu = 1,994$

9.15

Låt X vara vikten i hg i ett slumpmässigt valt paket. Vi vet att $X \sim N(\mu = 5,2, \sigma^2 = 0,1^2)$.

a) $P(X < 5) = P\left(\frac{X-\mu}{\sigma} < \frac{5-5,2}{0,1}\right) = P(Z < -2) = \Phi(-2) \approx 0,0228$. Låt nu Y vara antal paket bland 100000 som har en vikt mindre än 5 hg. Då är $Y \sim bin(n = 100000, p = 0,0228)$. Vi söker $E(Y) = np = 100000 \cdot 0,0228 = 2280$.

b) Vi söker μ sådant att $P(X < 5) = 0,002$. Detta innebär att $P\left(\frac{X-\mu}{\sigma} < \frac{5-\mu}{0,1}\right) = 0,002$, dvs. $\Phi\left(\frac{5-\mu}{0,1}\right) = 0,002$. Eftersom $\Phi(-2,88) \approx 0,002$ måste $\frac{5-\mu}{0,1} \approx -2,88$, dvs, $\mu \approx 5,288$.

9.16

Låt X vara antal personer i urvalet som är bärare av antikroppar. Då är $X \sim bin(n, p = 0,40)$

a) $n = 6$. $E(X) = np = 6 \cdot 0,40 = 2,4$. $P(X \leq 3) = 0,8208$

b) $n = 60$. $E(X) = np = 60 \cdot 0,40 = 24$. $P(X \leq 30)$

$= P\left(\frac{X-E(X)}{\sqrt{V(X)}} \leq \frac{30+0,5-24}{\sqrt{60 \cdot 0,4 \cdot (1-0,4)}}\right) \approx P(Z \leq 1,71) = 0,9564$

9.17

a) Låt X vara antal personer i urvalet som bor i villa. Då är $X \sim bin(n = 6, p = 0,60)$. Vi söker (i) $E(X) = np = 6 \cdot 0,60 = 3,6$ och (ii) $P(X \leq 3) = 0,4557$.

b) Låt Y vara antal personer i urvalet som bor i villa. Då är $Y \sim bin(n = 60, p = 0,60)$. Vi söker (i) $E(Y) = np = 60 \cdot 0,60 = 36$ och (ii) $P(Y \leq 30)$. Eftersom $np = 60 \cdot 0,60 = 36 > 5$ och $n(1-p) = 60 \cdot 0,40 = 24 > 5$ är approximation med normalfördelningen möjlig. Vi får att $P(Y \leq 30) = P(Y \leq 30,5) =$

$P\left(\frac{Y-np}{\sqrt{np(1-p)}} \leq \frac{30{,}5-60\cdot 0{,}6}{\sqrt{60\cdot 0{,}6\cdot(1-0{,}6)}}\right) \approx P(Z \leq -1,45) \approx 0,0735.$

9.18

Låt X vara arean i mm^2 hos ett slumpmässigt valt plaströr. Då är $X \sim N(\mu = 12,5, \sigma^2 = 0,2^2)$. Ett rör passar inte om arean är mindre än 12 mm^2 eller större än 13 mm^2 Sannolikheten att ett rör inte passar är

$p = P(X < 12 \text{ eller } X > 13) = 1-P(12 < X < 13) = 1-P\left(\frac{12-12{,}5}{0{,}2} < \frac{X-\mu}{\sigma} < \frac{13-12{,}5}{0{,}2}\right) = 1-P(-2,5 < Z < 2,5) = 1-(\Phi(2,5) - \Phi(-2,5)) = 1-(0,9938 - 0,0062) = 1 - 0,9876 = 0,0124$. Låt nu Y vara antal rör som inte passar i en låda med 1000 rör. Då är $Y \sim bin(n = 1000, p = 0,0124)$. Förväntat antal rör som inte passar är alltså $E(Y) = np = 1000 \cdot 0,0124 = 12,4$.

9.19

Låt X vara arean i mm^2 hos ett slumpmässigt valt plaströr. Då är $X \sim N(\mu = 12,5, \sigma^2)$. Om sannolikheten att arean överstiger 13 mm^2 är 0,05 måste service genomföras. Vi söker värdet på σ vid detta tillfälle. Det gäller att $P(X > 13) = 0,05$, dvs. $P\left(\frac{X-\mu}{\sigma} > \frac{13-12{,}5}{\sigma}\right) = 0,05$, $P\left(Z \leq \frac{0{,}5}{\sigma}\right) = 0,95$, $\Phi\left(\frac{0{,}5}{\sigma}\right) = 0,95$. Eftersom $\Phi(1,84) \approx 0,95$ får vi att $\frac{0{,}5}{\sigma} \approx 1,64$, dvs. $\sigma \approx 0,3049$.

9.20

a) Låt X vara antal kvinnor i styrelsen. Då är $X \sim bin(n = 10, p = 0,5)$. Det följer att $E(X) = np = 10 \cdot 0,5 = 5$ och $P(X < 4) = P(X \leq 3) = 0,1719$

b) Låt Y vara antal kvinnor i sångkören. Då är $Y \sim bin(n = 50, p = 0,5)$. Det följer att $E(Y) = np = 50 \cdot 0,5 = 25$. Vi söker $P(Y < 20)$ och undersöker om det är möjligt att använda normalfördelningsapproximation. Vi ser att $np = 50 \cdot 0,5 = 25 > 5$ och $n(1-p) = 50 \cdot 0,5 = 25 > 5$ dvs. det är möjligt att använda normalfördelningsapproximation. Vidare ser vi att $V(Y) = np(1-p) = 50 \cdot 0,5 \cdot 0,5 = 12,5$. Detta ger $P(Y < 20) = P\left(Y \leq 19 + \frac{1}{2}\right) \approx P\left(\frac{Y-E(Y)}{\sqrt{V(Y)}} \leq \frac{19+\frac{1}{2}-25}{\sqrt{12{,}5}}\right) \approx P(Z \leq -1,56) = \Phi(-1,56) \approx 0,0594$.

9.21

a) Låt X vara priset på aktien om ett halvår. Då är $X \sim N(\mu = 80, \sigma^2 = 4^2)$.

Vi söker $P(75 \leq X \leq 85) = P\left(\frac{75-80}{4} \leq \frac{X-\mu}{\sigma} \leq \frac{85-80}{4}\right) = P(-1,25 \leq Z \leq 1,25) = \Phi(1,25) - \Phi(-1,25) = 0,8944 - 0,1056 = 0,7888.$

b) Vi söker ett belopp k som är sådant att $P(X < k) = 0,01$, dvs. $P\left(\frac{X-\mu}{\sigma} < \frac{k-80}{4}\right) = 0,01$, $P\left(Z < \frac{k-80}{4}\right)$, $\Phi\left(\frac{k-80}{4}\right) = 0,01$. Eftersom $\Phi(-2,33) \approx 0,01$ får vi att $\frac{k-80}{4} \approx -2,33$, dvs. $k \approx 70,68$.

9.22

Låt X vara värdeförändringen från en dag till nästa. Då är $X \sim N(\mu = 0,6, \sigma^2 = 9)$.

a) $P(X > 0) = 1 - P(X \leq 0) = 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{0-0,6}{\sqrt{9}}\right) = 1 - P(Z \leq -0,2) = 1 - \Phi(-0,2) \approx 0,5793$

b) Sannolikheten att aktien minskar i värde från en dag till nästa är $p = 1 - 0,5793 = 0,4207$ Låt Y vara antal dagar under en vecka som aktien minskar i värde. Då är $Y \sim bin(n = 5, p = 0,4207)$. Vi söker $P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - 0,4207)^5 \approx 0,9348$

c) $5 \cdot 0,5793 = 2,896\,5$.

9.23

Låt X vara förbrukningen av etanol under en vecka. Då är $X \sim N(\mu = 50, \sigma^2 = 49)$.

a) Vi söker $P(X > 60) = 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{60-50}{7}\right) \approx 1 - P(Z \leq 1,43) = 1 - \Phi(1,43) \approx 1 - 0,9236 = 0,0764$.

b) Låt k vara den sökta lagerstorleken. Då gäller det att $P(X > k) = 0,01$, dvs. $P\left(\frac{X-\mu}{\sigma} \leq \frac{k-50}{7}\right) = 0,99$, $\Phi\left(\frac{k-50}{7}\right) = 0,99$. Eftersom $\Phi(2,33) \approx 0,99$ får vi att $\frac{k-50}{7} \approx 2,33$, dvs. $k = 66,31$

9.24

Låt A vara händelsen att få träff och X vara pulsfrekvensen. Vi vet att $P(A \mid X < 180) = 0,9$ och $P(A \mid X > 180) = 0,7$. Eftersom $X \sim N(\mu = 178, \sigma^2 = 4)$ följer det att $P(X < 180) = P\left(\frac{X-\mu}{\sigma} < \frac{180-178}{\sqrt{4}}\right) = P(Z < 1) = \Phi(1) \approx 0,8413$ och $P(X > 180) \approx 1 - 0,8413 = 0,1587$. Detta innebär att $P(A \cap (X < 180)) = P(X < 180)\, P(A \mid X < 180) \approx 0,8413 \cdot 0,9 = 0,75717$ och $P(A \cap (X > 180)) = P(X > 180)\, P(A \mid X > 180) \approx 0,1587 \cdot 0,7 = 0,11109$. Vi finner alltså sannolikheten att få en träff är $P(A) = P((A \cap (X < 180)) \cup (A \cap (X > 180))) =$

$P\left(A\cap(X<180)\right)+P\left(A\cap(X>180)\right)=0,75717+0,11109=0,86826$. Sannolikheten för fem träffar blir därför $0,86826^5\approx 0,4935$.

9.25

Låt X vara kursen vid börsens stängning. Vi vet att $X\sim N\left(\mu=82,\sigma^2=7,2\right)$.

a) Vi söker $P\left(X>86\right)=1-P\left(\frac{X-\mu}{\sigma}\leq\frac{86-82}{\sqrt{7,2}}\right)\approx 1-P\left(Z\leq 1,49\right)\approx 1-0,9319=0,0681$.

b) Låt Y vara antal dagar bland fem dagar som kursen överstiger 86 kr. Då är $Y\sim bin\left(n=5,p=0,0681\right)$. Vi söker $P\left(Y=3\right)=\begin{pmatrix}5\\3\end{pmatrix}0,0681^3\left(1-0,0681\right)^{5-2}\approx 0,0027$.

c) Vi söker $P\left(X<78\right)=P\left(\frac{X-\mu}{\sigma}<\frac{78-82}{\sqrt{7,2}}\right)\approx P\left(Z\leq -1,49\right)\approx 0,0681$.

d) Låt det sökta talet vara k. Då är $P\left(X<k\right)=0,10$, dvs. $P\left(\frac{X-\mu}{\sigma}<\frac{k-82}{\sqrt{7,2}}\right)=0,10$, $\Phi\left(\frac{k-82}{\sqrt{7,2}}\right)=0,10$. Eftersom $\Phi\left(-1,28\right)\approx 0,10$ får vi att $\frac{k-82}{\sqrt{7,2}}\approx -1,28$, $k\approx 78,57$.

9.26

a) Låt A vara händelsen att en slumpmässigt vald person är medlem i någon idrottsförening. Vi ser att $P\left(A\right)=\frac{51382}{128455}=0,4$

b) Låt X vara antal personer som är medlemmar i någon idrottsförening då 10 personer väljs slumpmässigt. Då är $X\sim bin\left(n=10,p=0,4\right)$. Det följer att $P\left(X=3,4 \text{ eller } 5\right)=P\left(3\leq X\leq 5\right)=P\left(X\leq 5\right)-P\left(X\leq 2\right)=0,8338-0,1673=0,6665$ och $P\left(0,3\leq\frac{X}{10}\leq 0,5\right)=P\left(3\leq X\leq 5\right)=0,6665$

c) Låt Y vara antal personer som är medlemmar i någon idrottsförening då 100 personer väljs slumpmässigt. Då är $Y\sim bin\left(n=100,p=0,4\right)$. (i) Vi söker $P\left(30\leq Y\leq 50\right)$ och undersöker om det är möjligt att använda normalfördelningsapproximation. Vi ser att $np=100\cdot 0,4=40>5$ och $n\left(1-p\right)=100\cdot\left(1-0,4\right)=60>5$, dvs. det är möjligt att använda normalfördelningsapproximation. Vidare ser vi att $E\left(Y\right)=np=100\cdot 0,4=40$ och $V\left(Y\right)=np\left(1-p\right)=100\cdot 0,4\cdot\left(1-0,4\right)=24$. Detta ger $P\left(30\leq Y\leq 50\right)=P\left(\frac{30-0,5-40}{\sqrt{24}}\leq\frac{Y-E(Y)}{\sqrt{V(Y)}}\leq\frac{50+0,5-40}{\sqrt{24}}\right)\approx P\left(-2,14\leq Z\leq 2,14\right)=$

$\Phi(2,14) - \Phi(-2,14) \approx 0,9838 - 0,0162 = 0,9676$. *(ii)* $P\left(0,3 \leq \frac{Y}{100} \leq 0,5\right) = P(30 \leq Y \leq 50) = 0,9676$

9.27

a) Låt X vara antal småhus i urvalet som har ett taxeringsvärde som överstiger 2800000 kr. Då är $X \sim bin(n = 400, p = 0,1587)$ (värdet på p ges i övning 9.5). Vi söker $P(X \geq 50)$. För det undersöker vi först om normalfördelningsapproximati kan användas. Det gäller att $np = 400 \cdot 0,1587 = 63,48 > 5$ och $n(1-p) = 400 \cdot (1 - 0,1587) = 336,52 > 5$, dvs. normalfördelningsapproximation kan användas. Eftersom $E(X) = np = 400 \cdot 0,1587 = 63,48$ och $V(X) = np(1-p) = 400 \cdot 0,1587 \cdot (1 - 0,1587) = 53,4057$ så att $P(X > 50) = P(X > 49,5) = P\left(\frac{X - E(X)}{\sqrt{V(X)}} > \frac{49,5 - 63,48}{\sqrt{53,4057}}\right) \approx P(Z > -1,91) = 1 - \Phi(-1,91) \approx 0,9719$.

b) Låt Y vara antal småhus i urvalet som har ett taxeringsvärde som överstiger 2800000 kr. Då är $Y \sim bin(n = 200, p = 0,1587)$ (värdet på p ges i övning 9.5). Vi söker $P(Y \geq 25)$. För det undersöker vi först om normalfördelningsapproximati kan användas. Det gäller att $np = 200 \cdot 0,1587 = 31,74 > 5$ och $n(1-p) = 200 \cdot (1 - 0,1587) = 168,26 > 5$, dvs. normalfördelningsapproximation kan användas. Eftersom $E(Y) = np = 200 \cdot 0,1587 = 31,74$ och $V(X) = np(1-p) = 200 \cdot 0,1587 \cdot (1 - 0,1587) = 26,7029$ så att $P(Y > 25) = P(Y > 24,5) = P\left(\frac{Y - E(Y)}{\sqrt{V(Y)}} > \frac{24,5 - 31,74}{\sqrt{26,7029}}\right) \approx P(Z > -1,40) = 1 - \Phi(-1,40) = 0,9192$.

9.28

Låt X vara klipptiden hos en slumpmässigt vald kund. Då är $X \sim N(\mu = 47, \sigma^2 = 6,5^2)$

a) Vi söker $P(X < 50) = P\left(\frac{X - \mu}{\sigma} < \frac{50 - 47}{6,5}\right) \approx P(Z < 0,4615) = \Phi(0,4615) \approx 0,6772$

b) Låt Y vara antal kunder med längre klipptid än 50 min. Sannolikheten att klipptiden är längre än 50 min är $1 - 0,6772 = 0,3228$, dvs. $Y \sim bin(n = 25, p = 0,3228)$. Vi söker $P(Y \geq 10)$ och undersöker om det är möjligt att använda normalfördelningsapproximation. Vi ser att $np = 25 \cdot 0,3228 = 8,07 > 5$ och $n(1-p) = 50 \cdot 0,6772 = 16,93 > 5$ dvs. det är möjligt att använda normalfördelningsapproximation. Vidare ser vi att $E(Y) = np = 25 \cdot 0,3228 = 8,07$ och $V(Y) = np(1-p) = 25 \cdot 0,3228 \cdot$

$0,6772 = 5,465$. Detta ger $P\left(Y \geq 10\right) = 1 - P\left(Y \leq 9\right) = 1 - P\left(\frac{Y - E(Y)}{\sqrt{V(Y)}} \leq \frac{9 + \frac{1}{2} - 8,07}{\sqrt{5,465}}\right) \approx 1 - P\left(Z \leq 0,6117\right) \approx 1 - 0,7291 = 0,2709$

c) Låt U vara antal kunder som köper en hårvårdsprodukt. Då är $U \sim bin\left(n = 25, p = 0,23\right)$.Vi söket $P\left(U \geq 1\right) = 1 - P\left(Y = 0\right) = 1 - \left(1 - 0,23\right)^{25} = 0,99855$.

9.29

a) Låt X vara tiden en person är arbetslös och att X har fördelningsfunktionen $F\left(x\right) = 1 - e^{-x/30}$. Vi söker $P\left(X \leq 69\right) = F\left(69\right) = 1 - e^{-69/30} \approx 0,9$

b) Låt Y vara antal arbetslösa efter 69 dagar i en grupp om 10 personer. Då är $Y \sim bin\left(n = 10, p = 0,9\right)$. Vi söker $P\left(Y \leq 1\right) = \left(1 - 0.9\right)^{10} + 10 \cdot 0,9 \cdot \left(1 - 0,9\right)^{9} \approx 0$

c) Låt nu Y vara antal arbetslösa efter 69 dagar i en grupp om 100 personer. Då är $Y \sim bin\left(n = 100, p = 0,9\right)$. Vi söker $P\left(Y \leq 10\right)$. Eftersom $np = 100 \cdot 0,9 = 90 > 5$ och $n\left(1 - p\right) = 100 \cdot 0,1 = 10 > 5$ är approximation med normalfördelning möjlig. Vi får att $P\left(Y \leq 10\right) = P\left(\frac{Y - np}{\sqrt{np(1-p)}} \leq \frac{10 + 0,5 - 100 \cdot 0,9}{\sqrt{100 \cdot 0,9(1 - 0,9)}}\right) \approx P\left(Z \leq -26,5\right) \approx 0$

23. SAMPLINGFÖRDELNINGAR OCH CENTRALA GRÄNSVÄRDESSATSEN

23.1

Proportionen sexor är en stokastisk variabel $P = \frac{X}{n}$, där X är antal sexor. Det gäller att $X \sim bin\left(n, p = \frac{1}{6}\right)$. Om P är i intervallet $\frac{1}{6} \pm \frac{1}{18}$, dvs. om $\frac{1}{9} \leq P \leq \frac{2}{9}$, så måste $\frac{1}{9}n \leq X \leq \frac{2}{9}n$.

a) Om $n = 20$ får vi $P\left(\frac{1}{9} \leq P \leq \frac{2}{9}\right) = P\left(\frac{20}{9} \leq X \leq \frac{40}{9}\right) = P(3 \leq X \leq 4) = 0,4401$.

b) Om $n = 40$ får vi $P\left(\frac{1}{9} \leq P \leq \frac{2}{9}\right) = P\left(\frac{40}{9} \leq X \leq \frac{80}{9}\right) = P(5 \leq X \leq 8) = 0,6068$.

c) Om $n = 80$ får vi $P\left(\frac{1}{9} \leq P \leq \frac{2}{9}\right) = P\left(\frac{80}{9} \leq X \leq \frac{160}{9}\right) = P(9 \leq X \leq 17) = 0,8128$.

23.2

Låt X vara vikten i gram av ett slumpmässigt valt pappersark. Då är $X \sim N(\mu = 80, \sigma = 4)$.

Låt Y vara vikten i gram av en bal med 5000 pappersark. Då är $E(Y) = \mu_Y = 5000 \cdot 80 = 400000$ och $V(Y) = \sigma_Y^2 = 5000 \cdot 4^2 = 80000$ så att $Y \sim N(\mu = 40000, \sigma^2 = 80000)$.

a) $P(Y < 399500) = P\left(\frac{Y-\mu_Y}{\sigma_Y} < \frac{399500-400000}{\sqrt{80000}}\right) = P(Z < -1,77) \approx 0,0384$, där $Z = \frac{Y-\mu_Y}{\sigma_Y} \sim N(0,1)$.

b) Låt U vara antal reklamerade balar. Då är $U \sim bin(n = 1000, p = 0,0384)$. Förväntat antal reklamationer blir $E(U) = np = 1000 \cdot 0,0384 = 38,4$.

c) En bal kostar $5000 \cdot 0,30 = 1500$ kr att tillverka. En ej reklamerad bal säljs för $5000 \cdot 0,40 = 2000$ kr. Vinst för en ej reklamerad bal är $2000 - 1500 = 500$ kr. En reklamerad bal säljs för 1200 kr. Vinst för en reklamerad bal är

$1200 - 1500 = -300$ kr. Förväntad vinst vid försäljning av en bal blir då $P\,(\text{ej reklamerad}) \cdot 500 + P\,(\text{reklamerad}) \cdot (-300) = (1 - 0,0384) \cdot 500 - 0,0384 \cdot 300 = 469,28$ kr.

d) Med de nya förutsättningarna kostar en bal att tillverka $5000 \cdot 0,28 = 1400$ kr. Vinst för ej reklamerad bal blir nu $2000 - 1400 = 600$ kr. Vinst för reklamerad bal är $1200 - 1400 = -200$ kr. Om genomsnittsvikten är 79,8 gram blir $E\,(Y) = \mu_Y = 5000 \cdot 79,8 = 399000$ så att sannolikheten för reklamation blir $P\,(Y < 399500) = P\left(\frac{Y - \mu_Y}{\sigma_Y} < \frac{399500 - 399000}{\sqrt{80000}}\right) = P\,(Z < 1,77) \approx 0,9616$ och förväntad vinst blir $P\,(\text{ej reklamerad}) \cdot 600 + P\,(\text{reklamerad}) \cdot (-200) = (1 - 0,9616) \cdot 600 - 0,9616 \cdot 200 = -169,28$ kr, dvs. det lönar sig inte att sänka vikten.

23.3

Låt X vara vikten i kg av en slumpmässigt vald person. Då är $X \sim N\,(\mu = 79, \sigma = 14)$.

Låt Y vara vikten i kg av 100 personer. Då är $E\,(Y) = \mu_Y = 100 \cdot 79 = 7900$ och $V\,(Y) = \sigma_Y^2 = 100 \cdot 14^2 = 19600$ så att $Y \sim N\,(\mu = 7900, \sigma^2 = 19600)$.

Vi söker $P\,(Y > 8200) = P\left(\frac{Y - \mu_Y}{\sigma_Y} > \frac{8200 - 7900}{\sqrt{19600}}\right) = P\,(Z > 2,14) \approx 0,0162$, där $Z = \frac{Y - \mu_Y}{\sigma_Y} \sim N\,(0,1)$.

23.4

Låt X vara IQ hos en slumpmässigt vald individ. Vi vet att $X \sim N\,(\mu = 100, \sigma = 15)$. Om $\overline{X}$ är urvalsmedelvärdet då n individer dras slumpmässigt är $\overline{X} \sim N\left(\mu = 100, \sigma_{\overline{X}} = \frac{15}{\sqrt{n}}\right)$ så att $P\left(95 \leq \overline{X} \leq 105\right) = P\left(\frac{95 - 100}{15/\sqrt{n}} \leq \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \leq \frac{105 - 100}{15/\sqrt{n}}\right) = P\left(-\frac{\sqrt{n}}{3} \leq Z \leq \frac{\sqrt{n}}{3}\right) = \Phi\left(\frac{\sqrt{n}}{3}\right) - \Phi\left(-\frac{\sqrt{n}}{3}\right)$.

(i) Om $n = 10$ får vi $P\left(95 \leq \overline{X} \leq 105\right) = \Phi\left(\frac{\sqrt{10}}{3}\right) - \Phi\left(-\frac{\sqrt{10}}{3}\right) \approx \Phi\,(1,05) - \Phi\,(-1,05) \approx 0,8531 - 0,1469 = 0,7062$.

(ii) Om $n = 100$ får vi $P\left(95 \leq \overline{X} \leq 105\right) = \Phi\left(\frac{\sqrt{100}}{3}\right) - \Phi\left(-\frac{\sqrt{100}}{3}\right) \approx \Phi\,(3,33) - \Phi\,(-3,33) \approx 0,9996 - 0,0004 = 0,9992$.

(iii) Om $n = 1000$ får vi $P\left(95 \leq \overline{X} \leq 105\right) = \Phi\left(\frac{\sqrt{1000}}{3}\right) - \Phi\left(-\frac{\sqrt{1000}}{3}\right) \approx \Phi\,(10,54) - \Phi\,(-10,54) \approx 1 - 0 = 1$.

23.5

Låt X vara taxeringsvärdet i kr för ett slumpmässigt valt småhus. Då är $X \sim N(\mu = 2800000, \sigma = 400000)$ och $V(X) = 400000^2$.

a) Låt $\overline{X}$ vara medelvärdet av taxeringsvärdena för $n = 25$ slumpmässigt valda småhus. Då är $E\left(\overline{X}\right) = \mu_{\overline{X}} = E(X) = 2800000$ kr, $V\left(\overline{X}\right) = \sigma_{\overline{X}}^2 = \frac{V(X)}{n} = \frac{400000^2}{25} = 6400000000$ så att standardavvikelsen är $\sigma_{\overline{X}} = \sqrt{6400000000} = 80000$. $P\left(2700000 < \overline{X} < 2900000\right) =$

$P\left(\frac{2700000-2800000}{80000} < \frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{2900000-2800000}{80000}\right) = P(-1,25 < Z < 1,25) = 0,8944-0,1056 = 0,7888$.

b) Låt $\overline{X}$ vara medelvärdet av taxeringsvärdena för n slumpmässigt valda småhus. Då blir standardavvikelsen för medelvärdet av taxeringsvärdena $\sigma_{\overline{X}} = \frac{400000}{\sqrt{n}}$. Vi söker n så att $P\left(2700000 < \overline{X} < 2900000\right) \geq 0,95$. På grund av symmetri kan vi ta $P\left(\overline{X} < 2900000\right) \geq 0,975$. Standardisering ger $P\left(\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{2900000-2800000}{400000/\sqrt{n}}\right) \geq 0,975$ eller $P\left(Z < \frac{100000}{400000/\sqrt{n}}\right) \geq 0,975$ där $Z = \frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} \sim N(0,1)$. Men $P(Z < 1,96) = 0,975$ så att $\frac{100000}{400000/\sqrt{n}} \geq 1,96$ och $\sqrt{n} \geq \frac{400000}{100000} \cdot 1,96$ dvs. $n \geq 61,47$. Det behövs minst 62 observationer.

23.6

Låt X vara höjden av en slumpmässigt vald solros. Vi vet att $X \sim N(\mu = 190, \sigma = 10)$.

a) Låt $\overline{X}$ vara medelvärdet av $n = 16$ slumpmässigt valda solrosor. Då är $\overline{X} \sim N\left(\mu = 190, \sigma_{\overline{X}} = \frac{10}{\sqrt{16}} = 2,5\right)$. Vi söker $P\left(188 \leq \overline{X} \leq 192\right) =$

$P\left(\frac{188-190}{2,5} \leq \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} \leq \frac{192-190}{2,5}\right) = P(-0,8 \leq Z \leq 0,8) = \Phi(0,8) - \Phi(0,8) \approx 0,7881 - 0,2119 = 0,5762$.

b) Vi söker n så att $P\left(188 \leq \overline{X} \leq 192\right) \geq 0,8$, dvs. så att $P\left(\frac{188-190}{10/\sqrt{n}} \leq \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} \leq \frac{192-190}{10/\sqrt{n}}\right) \geq 0,8$. Men då måste $P\left(\frac{\overline{X}-\mu}{\sigma_{\overline{X}}} \leq \frac{192-190}{10/\sqrt{n}}\right) \geq 0,9$, dvs. $\Phi\left(\frac{192-190}{10/\sqrt{n}}\right) \geq 0,9$. Eftersom $\Phi(1,28) \approx 0,9$ följer det att $\frac{192-190}{10/\sqrt{n}} \geq 1,28$, dvs. $\sqrt{n} \geq 1,28\frac{10}{2}$ och vi får att n måste vara minst 41.

23.7

Låt X vara IQ hos studenterna på utbildningslinjen. Då är $X \sim N(\mu = 105, \sigma = 10)$. Låt $\overline{X}$ vara medelvärdet av IQ hos n stycken slumpmässigt valda studenter. Då är $\overline{X} \sim N\left(\mu = 100, \sigma_{\overline{X}} = \frac{10}{\sqrt{n}}\right)$ så att $P\left(\overline{X} \geq 106\right) = 1 - P\left(\frac{\overline{X}-\mu}{\sigma_{\overline{X}}} \leq \frac{106-105}{10/\sqrt{n}}\right) = 1 - \Phi\left(\frac{\sqrt{n}}{10}\right)$.

i) Om $n = 25$ får vi $P\left(\overline{X} \geq 106\right) = 1 - \Phi\left(\frac{\sqrt{25}}{10}\right) = 1 - \Phi(0,5) = 0,3085$.

ii) Om $n = 100$ får vi $P\left(\overline{X} \geq 106\right) = 1 - \Phi\left(\frac{\sqrt{100}}{10}\right) = 1 - \Phi(1) = 0,1587$.

iii) Om $n = 900$ får vi $P\left(\overline{X} \geq 106\right) = 1 - \Phi\left(\frac{\sqrt{900}}{10}\right) = 1 - \Phi(3) = 0,0013$.

23.8

Låt X vara dragstyrkan hos ett slumpmässigt valt papper. Då är $X \sim N(\mu = 750, \sigma^2 = 100)$.

Då är sannolikheten att ett papper har en dragstyrka som är mindre än $C = 737,2$ lika med $P(X < C) = P(X < 737,2) = P\left(\frac{X-\mu}{\sigma} < \frac{737,2-750}{\sqrt{100}}\right) = P(Z < -1,28) \approx 0,1$, där $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Låt Y vara antal prover bland $n = 10$ prover som har en dragstyrka som är mindre än C. Då är $Y \sim bin(n = 10, p = 0,1)$. Vi söker sannolikheten att en rulle klassificeras som låg kvalitet, $P(Y \geq 3) = 1 - P(Y \leq 2) \approx 1 - 0,93 = 0,07$.

23.9

Låt X_1, X_2, X_3 och X_4 vara vikten hos passagerarna och Y vara summan av vikterna. Vi vet att $X_i \sim N(\mu = 75, \sigma = 16)$ för $i = 1,2,3,4$. Det följer då att $E(Y) = E(X_1 + X_2 + X_3 + X_4) = E(X_1) + E(X_2) + E(X_3) + E(X_4) = 75 + 75 + 75 + 75 = 300$ och $V(Y) = V(X_1 + X_2 + X_3 + X_4) = V(X_1) + V(X_2) + V(X_3) + V(X_4) = 16^2 + 16^2 + 16^2 + 16^2 = 1024$. Vi söker $P(Y > 325) = P\left(\frac{Y-E(Y)}{\sqrt{V(Y)}} > \frac{325-300}{\sqrt{1024}}\right) \approx 1 - P(Z \leq 0,78) \approx 1 - 0,7823 = 0,2177$.

23.10

Låt X vara vikten i gram av en slumpmässigt vald tablett. Då är $X \sim N(\mu = 0,65, \sigma = 0,02)$.

a) $P(0,64 < X < 0,66) = P\left(\frac{0,64-0,65}{0,02} < \frac{X-\mu}{\sigma} < \frac{0,66-0,65}{0,02}\right) = P(-0,5 < Z < 0,5) = 0,383$, där $Z \sim N(0,1)$.

b) $n = 30$ tabletter vägs. Då är medelvikten

$\overline{X} \sim N\left(\mu_{\overline{X}} = 0,65, \sigma_{\overline{X}} = 0,02/\sqrt{30} = 0,00365\right)$ och $P\left(0,64 < \overline{X} < 0,66\right) = P\left(\frac{0,64-0,65}{0,00365} < \frac{X-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{0,66-0,65}{0,00365}\right) \approx P(-2,74 < Z < 2,74) \approx 0,994$ så att sannolikheten att medelvärdet hamnar utanför intervallet är c:a $1 - 0,994 = 0,006$.

c) n tabletter vägs. Vi söker n så att sannolikheten att medelvärdet hamnar utanför intervallet är högst 0,05, dvs. $P\left(0,64 < \overline{X} < 0,66\right) \geq 0,95$. Standardisering ger $P\left(\frac{0,64-0,65}{0,02/\sqrt{n}} < \frac{X-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{0,66-0,65}{0,02/\sqrt{n}}\right) \geq 0,95$. På grund av symmetri kan vi ta $P\left(Z < \frac{0,01}{0,02/\sqrt{n}}\right) \geq 0,975$. Men $P(Z < 1,96) = 0,975$ så att $\frac{0,01}{0,02/\sqrt{n}} \geq 1,96$, dvs. $\sqrt{n} \geq \frac{0,02}{0,01}1,96$ och $n \geq 15,37$. Man måste ta minst $n = 16$ tabletter.

23.11

Proportionen invånare som ser tv-serien är $\pi = 0,35$. Låt P vara proportionen i ett urval om n invånare som har sett serien.

a) För $n = 30$ ser vi att $n\pi = 30 \cdot 0,35 = 10,5 > 5$ och att $n(1-\pi) = 30(1-0,35) = 19,5 > 5$, dvs. vi kan approximera fördelningen får P med $N\left(\mu = \pi = 0,35, \sigma = \sqrt{\pi(1-\pi)/n} \approx 0,0871\right)$. Vi söker $P((P < 0,3) \cup (P > 0,4)) = 1 - P(0,3 \leq P \leq 0,4) = 1 - P\left(\frac{0,3-0,35}{0,0871} \leq \frac{P-\pi}{\sigma} \leq \frac{0,4-0,35}{0,0871}\right) \approx 1 - P(-0,5741 \leq Z \leq 0,5741)$
$1 - (\Phi(0,5741) - \Phi(-0,5741)) \approx 1 - (0,7157 - 0,2843) = 0,5686$.

b) Med urvalsstorleken n blir variansen för P lika med $\sigma^2 = \pi(1-\pi)/n = 0,35(1-0,35)/n = 0,2275/n$. Vi söker n sådant att $P((P < 0,3) \cup (P > 0,4)) < 0,05$, dvs. $P(0,3 \leq P \leq 0,4) \geq 0,95$. Detta innebär att

$$P\left(\frac{0,3-0,35}{\sqrt{0,2275/n}} \leq \frac{P-\pi}{\sigma} \leq \frac{0,4-0,35}{\sqrt{0,2275/n}}\right) \approx P(-0,1048\sqrt{n} \leq Z \leq 0,1048\sqrt{n}) \geq 0,95,$$

dvs. $\Phi(0,1048\sqrt{n}) - \Phi(-0,1048\sqrt{n}) \geq 0,95$. Eftersom $\Phi(1,96) \approx 0,975$ följer det att $0,1048\sqrt{n} \geq 1,96$, dvs. $n \geq \left(\frac{1,96}{0,1048}\right)^2 \approx 349,6$. Det krävs minst 350 observationer.

23.12 (obs, står som övning 23.11 i boken)

Vi vet att $X \sim N(\mu, \sigma^2 = 4)$. Ett medelvärde av n slumpmässigt valda observationer blir då $\overline{X} \sim N\left(\mu, \sigma^2_{\overline{X}} = \frac{4}{n}\right)$. Händelsen A definieras som $A = -0,1 < \overline{X} - \mu < 0,1$. Vi söker n sådant att $P(A) \geq 0,95$, dvs. $P\left(-0,1 < \overline{X} - \mu < 0,1\right) = P\left(-\frac{0,1}{2/\sqrt{n}} < \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} < \frac{0,1}{2/\sqrt{n}}\right) \geq 0,95$ vilket är ekvivalent med att $\Phi\left(\frac{0,1}{2/\sqrt{n}}\right) - \Phi\left(-\frac{0,1}{2/\sqrt{n}}\right) \geq 0,95$. Eftersom $\Phi(1,96) \approx 0,975$ följer det att $\frac{0,1}{2/\sqrt{n}} \geq 1,96$ dvs. $n \geq \left(\frac{1,96 \cdot 2}{0,1}\right)^2 = 1536,6$. Det krävs alltså minst 1537 observationer för att sannolikheten är minst 0,95 att urvalsmedelvärdet avviker från μ med högst 0,1.

Standardavvikelsen för X är $\sigma = 2$ så att 10 procent av standardavvikelsen är 0,2. Händelsen B definieras därför som $B = -0,2 < \overline{X} - \mu < 0,2$. Vi söker nu n sådant att $P(B) \geq 0,95$, dvs. $P\left(-0,2 < \overline{X} - \mu < 0,2\right) = P\left(-\frac{0,2}{2/\sqrt{n}} < \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} < \frac{0,2}{2/\sqrt{n}}\right) \geq 0,95$ vilket är ekvivalent med att $\Phi\left(\frac{0,2}{2/\sqrt{n}}\right) - \Phi\left(-\frac{0,2}{2/\sqrt{n}}\right) \geq 0,95$ och det följer att $\frac{0,2}{2/\sqrt{n}} \geq 1,96$ dvs. $n \geq \left(\frac{1,96 \cdot 2}{0,2}\right)^2 = 384,16$. Det krävs alltså minst 385 observationer för att sannolikheten är minst 0,95 att urvalsmedelvärdet avviker från μ med högst 10 procent av standardavvikelsen av X.

24. UPPSKATTNING AV MODELLPARAMETRAR

24.1

a) Observationerna antas vara oberoende och normalfördelade med samma väntevärde och samma varians.

b) Observationernas medelvärde är $\overline{x} = 5,7$ och observationernas varians är $s^2 = 8,779$. Ett 99-procentigt konfidensintervall får genom $\overline{x} \pm t_{0,995}(n-1)\frac{s}{\sqrt{n}}$. Här är n=24 och $t_{0,995}(23) = 2,807$ får ur tabell. Det sökta konfidensintervallet är därför $5,7 \pm 1,698$.

c) I ett 90 procentigt konfidensintervall används faktorn $t_{0,95}(n-1)$ i stället för $t_{0,995}(n-1)$. Den förra faktorn är mindre, varför intervallet blir kortare.

24.2

Låt X vara maxhastigheten för en bil. Vi vet att $V(X) = \sigma^2 = 30$. $n = 50$ oberoende observationer tas. Eftersom n är stort kan samplingfördelningen för urvalsmedelvärdet $\overline{X}$ approximeras med $\overline{X} \sim N\left(\mu, \sigma^2_{\overline{X}} = \frac{30}{50}\right)$. Eftersom $\sum_{i=1}^{50} x_i = 8168$ blir det observerade medelvärdet $\overline{x} = \frac{8168}{50} = 163,36$ vilket tas som uppskattning av μ. Ett 95-procentigt konfidensintervall fås genom $\overline{x} \pm z_{0,975}\frac{\sigma}{\sqrt{n}}$. I detta fall får vi konfidensintervallet $163,36 \pm 1,96\frac{\sqrt{30}}{\sqrt{50}}$ dvs. $163,36 \pm 1,52$.

24.3

Alla falska.

24.4

Alla sanna.

24.5

Vi kan betrakta den observerade proportionen $p = 0,06$ som en observation på en normalfördelad stokastisk variabel med variansen $p\left(1-p\right)/n = 0,06\left(1-0,06\right)/1224 \approx 0,00004608$. Ett 95-procentigt konfidensintervall för populationsproportionen blir därför $p \pm z_{0,975}\sqrt{p\left(1-p\right)/n}$. I detta fall får vi konfidensintervallet $0,06 \pm 1,96\sqrt{0,00004608}$, dvs, $0,06 \pm 0,013$.

24.6

a) Vi använder urvalsproportionen som uppskattning av populationsproportionen, dvs. $p = 128/200 = 0,64$.

b) Ett 95-procentigt konfidensintervall fås genom $p \pm z_{0,975}\sqrt{p\left(1-p\right)/n}$, dvs. $0,64 \pm 1,96\sqrt{0,64\left(1-0,64\right)/200}$. Intervallet blir $0,64 \pm 0,0665$. Observera att urvalsstorleken är så pass liten (200/100000=0,002) så att hänsyn till ändlighetskorrektion inte behöver tas.

c) Urvalsstorleken och populationsproportionen har betydelse för precisionen av uppskattningen av populationsproportionen.

24.7

Låt X vara vikten av en potatispåse. Vi vet att $X \sim N\left(\mu, \sigma = 0,12\right)$.

a) $n = 20$ oberoende observationer gav urvalsmedelvärdet $\overline{x} = 2,53$. Ett 95-procentigt konfidensintervall blir då $2,53 \pm 1,96 \cdot \frac{0,12}{\sqrt{20}}$, dvs. $2,53 \pm 0,0526$

b) Vi önskar att den statistiska felmarginalen ska vara högst $B = 0,03$. Då behövs minst $n = 1,96^2\frac{\sigma^2}{B^2} = 1,96^2\frac{0,12^2}{0,03^2} \approx 61,5$, dvs. det krävs minst 62 observationer.

24.8

a) Vi beräknar $\overline{x} = \left(1,05 + 0,97 + 1,02 + 1,01 + 0,96\right)/5 = 1,002$ och vi vet att $\sigma = 0,05$ så att $\sigma_{\overline{X}} = \frac{0,05}{\sqrt{5}} = 0,0224$

95 %-igt kfi: $1,002 \pm 1,96 \cdot 0,0224$, dvs. $1,002 \pm 0,0438$

99 %-igt kfi: $1,002 \pm 2,58 \cdot 0,0224$, dvs. $1,002 \pm 0,0577$

99,9 %-igt kfi: $1,002 \pm 3,29 \cdot 0,0224$, dvs. $1,002 \pm 0,0736$

b) Totala längden av ett 95 %-igt kfi ska vara högst 0,08, dvs. den statistiska felmarginalen är högst $B = 0,04$. Då behövs minst $n = 1,96^2 \frac{\sigma^2}{B^2} = 1,96^2 \frac{0,05^2}{0,04^2} = 6,0025$, dvs. det krävs minst 7 observationer.

24.9

a) Låt $\overline{x}$ vara medelfördröjningstiden för industriföretag. Vi har att $n_x = 250$, $\overline{x} = 68,04$ och $s_x = 35,72$. Ett 95 %-igt kfi blir därför $68,04 \pm 1,96 \frac{35,72}{\sqrt{250}}$, dvs. $68,04 \pm 4,43$

b) Låt $\overline{y}$ vara medelfördröjningstiden för finansföretag. Vi har att $n_y = 238$, $\overline{y} = 56,74$ och $s_y = 34,87$. Ett 95 %-igt kfi blir därför $56,74 \pm 1,96 \frac{34,87}{\sqrt{238}}$, dvs. $56,74 \pm 4,43$

c) Det finns belägg för att tidsfördröjningarna är olika för industriföretag och finansföretag eftersom konfidensintervallen inte har några gemensamma punkter,

24.10

Vi kan anta att populationen är så stor att man kan bortse från ändlighetskorrektion. Urvalsstorleken är $n = 500$.

a) Medelbetalningsviljan uppskattas med urvalsmedelvärdet $\overline{x} = \frac{61000}{500} = 122$ kr.

b) Urvalsvariansen blir $s^2 = \frac{1}{500-1}\left(8692000 - \frac{61000^2}{500}\right) \approx 2505$. Standardavvikelsen av betalningsviljan uppskattas med urvalsstandardavvikelsen $s = \sqrt{2505} \approx 50$ kr.

c) Ett 95-procentigt konfidensintervall ges av $122 \pm 1,96 \cdot \frac{50}{\sqrt{500}}$, dvs. $122 \pm 4,383$.

d) Vi önskar att bredden av ett 95-procentigt konfidensintervall ska vara högst 20. Då blir den statistiska felmarginalen $B = 20/2 = 10$. Med standardavvikelsen 50 krävs då $n = 1,96^2 \frac{50^2}{10^2} = 96,04$, dvs. det krävs 97 observationer.

24.11

a) Medelvärdet av de uppmätta vikterna är $\overline{x} = \frac{4470}{9} = 496,67$ och variansen är $s^2 = \frac{1}{9-1}\left(2220404 - \frac{4470^2}{9}\right) = 38$, så att standardavvikelsen blir $\sqrt{38}$.

Ett 95-procentigt konfidensintervall fås genom $496,67 \pm 2,306 \cdot \frac{\sqrt{38}}{\sqrt{9}}$, dvs. $496,67 \pm 4,73$, där 2,306 är 97,5-percentilen i t-fördelningen med $9-1=8$ frihetsgrader.

b) Antag nu at variansen är $\sigma^2 = 36$. Vi kan därför använda normalfördelningen som samplingfördelning. Vi önskar att den statistiska felmarginalen ska vara högst 2 vid konfidensgraden 99 procent. Det krävs då minst $n = 2,58^2 \frac{36}{2^2} = 59,9076$, dvs. det krävs minst 60 observationer.

24.12

Vi antar att energiförbrukningen är normalfördelad och att observationerna är oberoende. Ett 95-procentigt konfidensintervall fås genom $45357 \pm 2,064 \frac{7213}{\sqrt{25}}$, dvs. $45357 \pm 2977,5$.

24.13

a) Populationsproportionen kan uppskattas med urvalsproportionen $p = \frac{122}{500} = 0,244$, dvs. man kan uppskatta att 24,4 procent av kommunens invånare är beredda att betala 100 kr per år i miljöavgift.

b) Variansen av urvalsproportionen uppskattas med $\frac{0,244(1-0,244)}{500} = 0,000368928$ så att standardavvikelsen uppskattas med $\sqrt{0,000368928} \approx 0,019$.

c) Ett 95-procentigt konfidensintervall ges av $0,244 \pm 1,96 \cdot 0,019$, dvs. $0,244 \pm 0,0376$.

d) Vi önskar att den statistiska felmarginalen är högst $0,1/2 = 0,05$. Det krävs då $n = 1,96^2 \frac{0,244(1-0,244)}{0,05^2} \approx 283,45$, dvs. det krävs minst 284 observationer.

24.14

a) Medellivslängden av de testade däcken är $\overline{x} = \frac{302,4}{6} = 50,4$. Ett 95-procentigt konfidensintervall blir därför $50,4 \pm 1,96 \frac{5,5}{\sqrt{6}}$, dvs. $50,4 \pm 4,4$.

b) Vi önskar en statistisk felmarginal om högst $6/2 = 3$. Det krävs då minst $n = 1,96^2 \frac{5,5^2}{3^3} \approx 12,91$, Dvs. det krävs minst 13 observationer.

24.15

a) Populationsproportionen kan uppskattas med urvalsproportionen $p = \frac{68}{400} = 0,17$.

b) Urvalet är mindre än 5 procent av populationen. Vi behöver därför inte använda ändlighetskorrektion. Variansen av proportionen i a) uppskattas med $\frac{0{,}17(1-0{,}17)}{400} = 0,00035275$.

c) Ett 95-procentigt konfidensintervall ges av $0,17 \pm 1,96\sqrt{0,00035275}$, dvs. $0,17 \pm 0,037$.

d) C:a 95 procent av alla 95-procentiga konfidensintervall som kan göras kommer att innehålla populationsmedelvärdet. Huruvida det specifika konfidensintervallet som beräknades i uppgift c) innehåller populationsmedelvärdet eller ej vet vi inte.

e) Vi önskar att den statistiska felmarginalen ska vara högst 0,03. Populationsproportionen är okänd men antas inte överstiga 25 procent. Variansen är då högst $0,25\,(1-0,25) = 0,1875$. Det krävs därför $n = 1,96^2 \frac{0{,}1875}{0.03^2} = 800,3$, dvs. det krävs minst 801 observationer.

25. URVAL FRÅN ÄNDLIGA POPULATIONER

25.1

a) Antal felaktigt bokförda fakturor är 11 bland 100 undersökta. En uppskattning av proportionen felaktigt bokförda är därför $\widehat{p} = \frac{11}{100} = 0,11$. En uppskattning av variansen för $\widehat{p}$ är $\frac{\widehat{p}(1-\widehat{p})}{n}\left(\frac{N-n}{N-1}\right) = \frac{0,11(1-0,11)}{100}\left(\frac{1100-100}{1100-1}\right) \approx 0,00089$. Ett 95 %-igt konfidensintervall är $0,11 \pm 1,96\sqrt{0,00089}$, dvs. $0,11 \pm 0,06$.

b) Sammanlagda bokföringsfelet är 22000 kr bland 100 undersökta fakturor. En uppskattning av bokföringsfelet i medeltal är därför $\overline{x} = \frac{22000}{100} = 220$ kr. Vi vet att urvalsstandardavvikelsen av bokföringsfelet är $s = 700$ kr. En uppskattning av populationsvariansen är då $\widehat{\sigma}^2 = s^2\left(\frac{N-n}{N-1}\right) = 700^2\left(\frac{1100-100}{1100-1}\right) = 700^2 \cdot 0,9099$. Ett 95 %-igt konfidensintervall för det genomsnittliga bokföringsfelet blir $220 \pm 1,96 \cdot \frac{700\cdot\sqrt{0,9099}}{\sqrt{100}}$, dvs. $220 \pm 130,87$.

25.2

Låt μ vara genomsnittligt antal fel. Vi uppskattar μ med urvalsmedelvärdet $\overline{x}$. Populationsvariansen är okänd och uppskattas urvalsvariansen s^2. Vi får att $\sum x = 35$, $\sum x^2 = 69$ så att $\overline{x} = 35/35 = 1$ och $s^2 = (69 - 35^2/35)\,/\,(35-1) = 1$. Med ändlighetskorrektion blir en uppskattning av variansen för urvalsmedelvärdet är därför $(1/35) \cdot (400-35)\,/\,(400-1) = 0,0261$. Eftersom $n > 30$ kan vi använda normalfördelningsapproximation. . Detta ger intervallet $1 \pm 1.96 \cdot \sqrt{0,0261}$, vilket ger konfidensintervallet $1 \pm 0,32$.

25.3

Proportionen felaktiga adresser uppskattas till $80/400 = 0,2$ med variansen $0,2\,(1-0,2)\,/\,(400-1)\,((3000-400)\,/\,(3000-1)) = 0,00026071$. Ett konfidensintervall för proportionen blir därför $0,2 \pm 1,96\sqrt{0,00026071}$ dvs. $0,2 \pm 0,016$. Multiplicera med populationsstorleken får vi ett intervall för antalet felaktiga adresser: 600 ± 48 (avrundat till heltal).

25.4

Den statistiska felmarginalen är $B = 0,7$ cm och populationsstandardavvikelsen antas vara $\sigma = 5$ cm. Vi får då $n_0 = 1,96^2 \cdot 5^2/0,7^2 = 196$ och $n = 196/\left(1 + 196/1000\right) = 163,9$ dvs. det krävs minst 164 observationer.

25.5

a) Vi vet att $N = 600$, $n = 200$ så att $\overline{x} = 1373/200 = 66,865$. Vidare vet vi att $s = 264$. Ett kfi blir då $66,865 \pm 1,96 \cdot \frac{264}{\sqrt{200}}\left(\frac{600-200}{600-1}\right)$, dvs. $66,865 \pm 24,433$

b) Vi antar att den utvidgade populationen är så stor att vi kan bortse från ändlighetskorrektion. Vi sätter den statistiska felmarginalen till $B = 24,433$ och får $n = 1,96^2 \cdot 264^2/24,433^2 = 448,5$ dvs. det behövs 449 observationer.

25.6

a) Vi kan anta att populationen är så stor att man kan bortse från ändlighetskorrektion. Vi vet att urvalsmedelvärdet är 2236200 kr och urvalsstandardavvikelsen är 346000 kr. Ett 95-procentigt konfidensintervall blir då $2236200 \pm 1,96\frac{346000}{\sqrt{300}}$, dvs. 2236200 ± 39154.

b) Populationsstorleken är nu $N = 75$ och vi önskar en statistisk felmarginal som är högst $B = 39154/2 = 19577$. Det krävs då minst $n = \frac{346000^2}{\frac{19577^2}{4} + \frac{346000^2}{75}} = 70,75$, det krävs minst 71 observationer.

25.7

a) Medelvärdet av de nio kvittonas summor är $\frac{1060}{16} = 66,25$ med variansen $\frac{1}{16-1}\left(78810 - \frac{1060^2}{16}\right) = 572,33$. En uppskattning av urvalsmedelvärdets varians blir därmed $\frac{572,33}{16}\left(\frac{140-16}{140-1}\right) = 31,91$. En uppskattning av totala dagsförsäljningen fås genom att multiplicera medelvärdet per kassakvitto med antalet kvitton, dvs. vi får uppskattningen som $140 \cdot 66,25 = 9275$ med variansen $140^2 \cdot 31,91 = 625436$. Ett 95-procenigt konfidensintervall för totala dagsförsäljningen blir därför $9275 \pm 2,131 \cdot \sqrt{625436}$, dvs. 9275 ± 1685, där 2,131 är 97,5-percentilen i t-fördelningen med $16 - 1 = 15$ frihetsgrader.

b) Antag nu att variansen av summorna på kvittona är 900. Variansen för urvalsmedelvärdet vid n observationer blir då $900/n\left(\frac{140-n}{140-1}\right)$ och variansen för uppskattningen av totalen blir $140^2 \cdot 900/n\left(\frac{140-n}{140-1}\right)$. Vi önskar en statistisk felmarginal om högst 1000 kr vid 99-procentig konfidensgrad. Detta innebär att $2,58 \cdot \sqrt{140^2 \cdot 900/n \left(\frac{140-n}{140-1}\right)} \leq 500$, dvs. $2,58^2 \cdot 140^2 \cdot 900 \cdot 140 \leq (500^2 \cdot (140-1) + 2,58^2 \cdot 140^2 \cdot 900) \cdot n$, $n \geq \frac{2,58^2 \cdot 140^2 \cdot 900 \cdot 140}{500^2 \cdot (140-1) + 2,58^2 \cdot 140^2 \cdot 900} \approx 108,03$, dvs. det krävs minst 109 observationer.

25.8

a) Förväntat värde av summan på ett kvitto kan uppskattas med urvalsmedelvärdet $\overline{x} = \frac{1147}{8} = 143,38$.

b) Vi har ett urval om $n = 8$ kvitton ur en population om $N = 44$ kvitton. Ändlighetskorrektion måste tillämpas eftersom urvalet är större än 5 procent av populationen. Urvalsvariansen är $s^2 = \frac{1}{8-1}\left(200331 - \frac{1147^2}{8}\right) = 5125,70$, så att variansen för medelvärdet är $\frac{5125,70}{8}\frac{44-8}{44-1} = 591,43$ och standardavvikelsen $\sqrt{591,43} = 24,32$

c) Ett 95-procentigt konfidensintervall fås genom $143,38 \pm 2,365 \cdot 24,32$, dvs. $143,38 \pm 57,52$, där 2,365 är 97,5-percentilen i t-fördelningen med $8 - 1 = 7$ frihetsgrader.

d) C:a 95 procent av alla 95-procentiga konfidensintervall som kan göras kommer att innehålla populationsmedelvärdet. Huruvida det specifika konfidensintervallet som beräknades i uppgift c) innehåller populationsmedelvärdet eller ej vet vi inte.

25.9

Populationsstorleken är $N = 1775$ och den statistiska felmarginalen är $B = 0,06/2 = 0,03$. Den största variansen vi kan ha är $\sigma^2 = 0,25$. Minsta urvalsstorlek blir då $n = \frac{0,25}{\frac{0,03^2}{4} + \frac{0,25}{1775}} = 683,35$. Det behövs minst 684 observationer.

25.10

Populationsstorleken är $N = 2500$ och den statistiska felmarginalen är $B = 0,10/2 = 0,05$. Den största variansen vi kan ha är $\sigma^2 = 0,25$. Minsta urvalsstorlek blir då $n = \frac{0,25}{\frac{0,05^2}{4} + \frac{0,25}{2500}} = 344,82$. Det behövs minst 345 observationer.

25.13

a) Vi vet att $N_N = 890$, $N_C = 2200$ och $N_S = 1170$ så att $N = 890 + 2200 + 1170 = 4260$. Detta ger oss $w_N = 890/4260 = 0,2089$, $w_C = 2200/4260 = 0,5164$ samt $w_S = 1170/4260 = 0,2747$. Vidare vet vi att $\sigma_N = 40000$, $\sigma_C = 30000$ och $\sigma_S = 20000$. Vi önskar en statistisk felmarginal om högst $B = 2000$. Detta ger en total urvalsstorlek om $n = 1,96^2 \frac{(0,2089 \cdot 40000 + 0,5164 \cdot 30000 + 0,2747 \cdot 20000)^2}{2000^2} = 826,86$ dvs. det behövs minst $n = 827$ observationer.

b) $n_N = 827 \frac{0,2089 \cdot 40000}{0,2089 \cdot 40000 + 0,5164 \cdot 30000 + 0,2747 \cdot 20000} = 235,51$,

$n_C = 827 \frac{0,5164 \cdot 30000}{0,2089 \cdot 40000 + 0,5164 \cdot 30000 + 0,2747 \cdot 20000} = 436,64$

$n_S = 827 \frac{0,2747 \cdot 20000}{0,2089 \cdot 40000 + 0,5164 \cdot 30000 + 0,2747 \cdot 20000} = 154,85$.

Vi avrundar uppåt till hela tal, $n_N = 236$, $n_C = 437$ och $n_S = 155$.

c) Vi vet att $\overline{x}_N = 95000$, $\overline{x}_C = 65000$ och $\overline{x}_S = 50000$. En uppskattning av medelinkomsten blir då $\overline{x} = 0,2089 \cdot 95000 + 0,5164 \cdot 65000 + 0,2747 \cdot 50000 = 67147$.

d) Variansen för det stratifierade medelvärdet är $0,2089^2 \frac{40000^2}{236} \left(\frac{890-236}{890-1}\right) + 0,5164^2 \frac{30000^2}{437} \left(\frac{2200-437}{2200-1}\right) + 0,2747^2 \frac{20000^2}{155} \left(\frac{1170-155}{1170-1}\right) \approx 827050$. Ett 95 procentigt kfi för medelinkomsten blir därmed $67147 \pm 1,96\sqrt{827050}$, dvs $67147 \pm 1782,5$

25.14

a) Totala urvalsstorleken är $n = 1000$. Med proportionell allokering får vi urvalsstorlekarna i varje stratum som $n_1 = \frac{10000}{25000} 1000 = 400$, $n_2 = \frac{10000}{25000} 1000 = 400$ och $n_3 = \frac{5000}{25000} 1000 = 200$. Uppskattningen $\overline{X}_{prop} = \frac{10000}{25000} \overline{X}_1 + \frac{10000}{25000} \overline{X}_2 + \frac{5000}{25000} \overline{X}_3 = 0,4\overline{X}_1 + 0,4\overline{X}_2 + 0,2\overline{X}_3$ får variansen $V\left(\overline{X}_{prop}\right) = 0.4^2 \frac{900}{400} \left(\frac{10000-400}{10000-1}\right) + 0.4^2 \frac{1600}{400} \left(\frac{10000-400}{10000-1}\right) + 0.2^2 \frac{6400}{200} \left(\frac{5000-200}{5000-1}\right) = 2.189$

b) Vid Neymanallokering får vi $n_1 = 1000 \frac{0.4\sqrt{900}}{0.4\sqrt{900} + 0.4\sqrt{1600} + 0.2\sqrt{6400}} \approx 273$, $n_2 = 1000 \frac{0.4\sqrt{1600}}{0.4\sqrt{900} + 0.4\sqrt{1600} + 0.2\sqrt{6400}} \approx 364$ och $n_3 = 1000 \frac{0.2\sqrt{6400}}{0.4\sqrt{900} + 0.4\sqrt{1600} + 0.2\sqrt{6400}} \approx 363$. Urvalsstorlekarna är avrundade till heltal så att de summerar till 1000. Variansen för det stratifereade medelvärdet blir $0.4^2 \frac{900}{273} \left(\frac{10000-273}{10000-1}\right) + 0.4^2 \frac{1600}{364} \left(\frac{10000-364}{10000-1}\right) + 0.2^2 \frac{6400}{363} \left(\frac{5000-364}{5000-1}\right) = 1,844$.

25.15

a) Vi vet att $N_1 = 200$, $N_2 = 300$ och $N_3 = 500$ så att $N = 200+300+500 = 1000$ och $w_1 = 200/1000 = 0,2$, $w_2 = 300/1000 = 0,3$ samt $w_3 = 500/1000 = 0,5$. Vidare vet vi att $\sigma_1 = \sqrt{900} = 30$, $\sigma_2 = \sqrt{625} = 25$ och $\sigma_3 = \sqrt{81} = 9$. Med $n = 240$ ger detta Neymanallokeringen $n_1 = 240\frac{0{,}2\cdot 30}{0{,}2\cdot 30+0{,}3\cdot 25+0{,}5\cdot 9} = 80$, $n_2 = 240\frac{0{,}3\cdot 25}{0{,}2\cdot 30+0{,}3\cdot 25+0{,}5\cdot 9} = 100$ och $n_3 = 240\frac{0{,}5\cdot 9}{0{,}2\cdot 30+0{,}3\cdot 25+0{,}5\cdot 9} = 60$.

b) En väntevärdesriktig uppskattning av populationsmedelvärdet fås genom ett vägt medelvärde av stratummedelvärdena och med stratumproportionerna som vikter. $\overline{x}_{st} = 0,2 \cdot 480 + 0,3 \cdot 360 + 0,5 \cdot 180 = 294$

c) Vi önskar en varians som inte överstiger 4. Detta innebär att $\frac{(0{,}2\cdot 30+0{,}3\cdot 25+0{,}5\cdot 9)^2}{n} = \frac{324}{n} \leq 4$, dvs. det behövs minst $n \geq \frac{324}{4} = 81$ observationer.

25.16

a) En uppskattning av proportionen elever som är positiva till skolans salladsbord är $\frac{350}{900} \cdot \frac{48}{60} + \frac{300}{900} \cdot \frac{25}{50} + \frac{250}{900} \cdot \frac{8}{40} = \frac{350}{900} \cdot 0.8 + \frac{300}{900} \cdot 0.5 + \frac{250}{900} \cdot 0.2 = 0.53333$. När vi uppskattar variansen av uppskattningen måste vi tänka på att vi bara har tillgång till uppskattningar av stratumvarianserna. Variansen för uppskattningen uppskattas därför med $\left(\frac{350}{900}\right)^2 \frac{0.8\cdot(1-0.8)}{60} \left(\frac{350-60}{350}\right) + \left(\frac{300}{900}\right)^2 \frac{0.5\cdot(1-0.5)}{50} \left(\frac{300-50}{300}\right) + \left(\frac{250}{900}\right)^2 \frac{0.2\cdot(1-0.2)}{40} \left(\frac{250-40}{250}\right) = 0.001056$ så att ett 95 %-igt konfidensintervall är $0.53333 \pm 1.96\sqrt{0.001056}$ dvs. $0.53333 \pm 0,06369$.

b) Vi använder Neymanallokering. $n_k = n\frac{N_k\sigma_k}{N_1\sigma_1+N_2\sigma_2+N_3\sigma_3}$ där $n = 150$ och vi använder uppskattade standardavvikelser för σ. Vi ser att $N_1\sigma_1 + N_2\sigma_2 + N_3\sigma_3 = 350\sqrt{0.8 \cdot 0.2} + 300\sqrt{0.5 \cdot 0.5} + 250\sqrt{0.2 \cdot 0.8} = 390$ Detta ger $n_1 = 150 \cdot 350\sqrt{0.8 \cdot 0.2}/390 = 53.846$, $n_2 = 150 \cdot 300\sqrt{0.5 \cdot 0.5}/390 = 57.692$ och $n_2 = 150 \cdot 250\sqrt{0.2 \cdot 0.8}/390 = 38.462$. Avrunda till $n_1 = 54$, $n_2 = 58$ och $n_3 = 38$.

25.17

$E\left(\widehat{\pi}\left(1-\widehat{\pi}\right)\right) = E\left(\frac{X}{n}\left(1-\frac{X}{n}\right)\right) = E\left(\frac{X}{n} - \frac{X^2}{n^1}\right) = E\left(\frac{X}{n}\right) - E\left(\frac{X^2}{n^1}\right) = \frac{E(X)}{n} - \frac{E\left(X^2\right)}{n^2}$. Vi vet att $E\left(X\right) = n\pi$ och $V\left(X\right) = n\pi\left(1-\pi\right)\frac{N-n}{N-1}$. Eftersom $V\left(X\right) = E\left(X^2\right) - \left(E\left(X\right)\right)^2$ följer det att $E\left(X^2\right) = V\left(X\right) + \left(E\left(X\right)\right)^2 = n\pi\left(1-\pi\right)\frac{N-n}{N-1} + \left(n\pi\right)^2$. Detta ger $E\left(\frac{X}{n}\left(1-\frac{X}{n}\right)\right) = \frac{n\pi}{n} - \frac{n\pi(1-\pi)\frac{N-n}{N-1}+(n\pi)^2}{n^2} =$

$$\pi\left(1-\pi-(1-\pi)\frac{N-n}{(N-1)n}\right)=\pi(1-\pi)\left(1-\frac{N-n}{(N-1)n}\right)=\pi(1-\pi)\left(\frac{(N-1)n-(N-n)}{(N-1)n}\right)=$$
$$\pi(1-\pi)\left(\frac{Nn-N}{(N-1)n}\right)=\pi(1-\pi)\left(\frac{N(n-1)}{(N-1)n}\right).$$

26. PRÖVNING AV HYPOTESER OM MODELLPARAMETRAR

26.1

Låt X vara vikten av en förpackning, $X \sim N(\mu, \sigma = 0,10)$ och $\overline{X}$ är medelvikten av $n = 12$ oberoende observationer på vikten.

$H_0 : \mu = 2,0$

$H_A : \mu \neq 2,0$

Signifikansnivå $\alpha = 0,05$

Testvariabel $Z = \frac{\overline{X}-2{,}0}{0{,}10/\sqrt{12}}$

Om H_0 är sann är $Z \sim N(0,1)$

H_0 förkastas om $|z_{obs}| > 1,96$

$z_{obs} = \frac{1{,}95-2{,}0}{0{,}10/\sqrt{12}} = -1,73$, dvs. H_0 kan inte förkastas på 5 %-nivån. Det finns inte stöd för att tro att förpackningarna inte har en genomsnittsvikt av 2 kg.

26.2

Låt π vara proportionen felaktiga förpackningar. Vi vill pröva

$H_0 : \pi = 0,08$

$H_A : \pi \neq 0,08$ (Om man vill pröva om felfrekvensen har minskat blir $H_A : \pi < 0,08$)

Låt P vara proportionen felaktiga förpackningar i urvalet. Om H_0 är sann är P approximativt $N\left(\pi = 0.08, \frac{\pi(1-\pi)}{160} = \frac{0{,}08(1-0{,}08)}{160} = 0,00046\right)$.

Signifikansnivån är $\alpha = 0,05$

Vi använder $Z = \frac{P-0{,}08}{\sqrt{0{,}00046}}$ som testvariabel och förkastar H_0 om observerat värde på Z är mindre än -1,96 eller större än 1,96.

10 förpackningar var felaktiga, varför $p = \frac{10}{160} = 0,0625$ och $z_{obs} = \frac{0,0625-0,08}{\sqrt{0,00046}} \approx -0,82$. Eftersom $|z_{obs}| < 1,96$ kan H_0 inte förkastas, dvs. vi kan inte påvisa att åtgärderna har haft någon effekt på den genomsnittliga felfrekvensen.

26.3

Låt X vara livslängden hos ett skärverktyg, $X \sim N(\mu, \sigma)$ och $\overline{X}$ är medellivslängden av $n = 6$ oberoende observationer på livslängden.

$H_0 : \mu = 3000$

$H_A : \mu < 3000$

Signifikansnivå $\alpha = 0,05$

Testvariabel $t = \frac{\overline{X}-3000}{s/\sqrt{6}}$

Om H_0 är sann är $t \sim t(5)$

H_0 förkastas om $t_{obs} < t_{0,05}(5) = -2,015$

$\sum x_i = 2970 + 3020 + 3005 + 2900 + 2940 + 2925 = 17760$

$\sum x_i^2 = 2970^2 + 3020^2 + 3005^2 + 2900^2 + 2940^2 + 2925^2 = 52580550$

$\overline{x} = \sum x_i/n = 17760/6 = 2960$

$s^2 = \left(\sum x_i^2 - \left(\sum x_i\right)^2/n\right)/(n-1) = \left(52580550 - 17760^2/6\right)/(6-1) = 2190$

$t_{obs} = \frac{2960-3000}{\sqrt{2190}/\sqrt{6}} = -2,09$, dvs. H_0 kan förkastas på 5 %-nivån. Det finns stöd för att tro att livslängden är mindre än 3000 skär.

26.4

Låt π vara proportionen individer som känner till produkten. Vi vill pröva

$H_0 : \pi = 0,40$

$H_A : \pi < 0,40$

Låt P vara proportionen i urvalet som känner till produktnamnet. Om H_0 är sann är P approximativt $N\left(\pi = 0,40; \frac{\pi(1-\pi)}{500} = \frac{0,40(1-0,40)}{500} = 0,0,00048\right)$.

Vi väljer signifikansnivån $\alpha = 0,05$ och testvariabeln $Z = \frac{P-0,40}{\sqrt{0,00048}}$. Vi förkastar H_0 om observerat värde på Z är mindre än -1,64.

180 personer kände till produktnamnet, dvs. $p = \frac{180}{500} = 0,36$ och $z_{obs} = \frac{0,36-0,40}{\sqrt{0,00048}} \approx -1,826$. Eftersom $z_{obs} < -1,64$ kan H_0 förkastas på 5-procentsnivån, dvs. det finns belägg för att mindre än 40 procent av kunderna på marknaden inte känner till produktnamnet.

26.5

Låt X vara antal livslängden hos en glödlampa med $E(X) = \mu$ och $\overline{X}$ är medellivslängden av $n = 30$ oberoende observationer på livslängden.

$H_0 : \mu = 800$

$H_A : \mu \neq 800$

Signifikansnivå $\alpha = 0,01$

Testvariabel $t = \frac{\overline{X}-800}{S/\sqrt{30}}$

Om H_0 är sann är t approximativt $N(0,1)$ (eftersom $n \geq 30$)

H_0 förkastas om $|t_{obs}| > 2,5758$

$\sum x_i = 19640$

$\sum x_i^2 = 13468200$

$\overline{x} = \sum x_i / n = 19640/30 = 654,67$

$s^2 = \left(\sum x_i^2 - \left(\sum x_i\right)^2 / n\right) / (n-1) = \left(13468200 - 19640^2/30\right) / (30-1) = 21053,33$

$t_{obs} = \frac{654,67-800}{\sqrt{21053,33}/\sqrt{30}} = -5,49$, dvs. H_0 kan förkastas på 1 %-nivån. Det finns stöd för att tro att livslängden inte är mindre 800 timmar.

26.6

Låt X vara antal livslängden hos en komponent med $E(X) = \mu$ och standardavvikelsen $\sigma = 300$. $\overline{X}$ är medellivslängden av $n = 100$ oberoende observationer på livslängden. Vi vill pröva

$H_0 : \mu = 5200$

$H_A : \mu > 5200$

på signifikansnivån $\alpha = 0,05$

med testvariabeln $t = \frac{\overline{X}-5200}{\sigma/\sqrt{100}}$

Om H_0 är sann är $t \sim N(0,1)$ (eftersom $n \geq 30$)

H_0 förkastas om $|t_{obs}| > 1,64$

Observerat medelvärde i urvalet är $\overline{x} = 5265$ så att $t_{obs} = \frac{5265-5200}{300/\sqrt{100}} = 2,167$, dvs. H_0 kan förkastas på 5 %-nivån. Det finns stöd för att tro att den nya produktionsprocessen ger en längre livslängd än den gamla.

26.7

Låt P vara urvalsproportionen sympatisörer. I stora urval (stort n) är $P \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$, approximativt, där π är populationsproportionen sympatisörer.

$H_0 : \pi = 0,20$

$H_A : \pi > 0,20$

Signifikansnivå $\alpha = 0,05$

Testvariabel $Z = \frac{P-0,2}{\sqrt{\frac{0,2(1-0,2)}{400}}} = \frac{P-0,2}{0,02}$

Om H_0 är sann är Z approximativt $N(0,1)$ (eftersom $n \geq 30$)

H_0 förkastas om $z_{obs} > 1,64$

Urvalsproportionen observerades till $p = 92/400 = 0,23$

$z_{obs} = \frac{0,23-0,20}{0,02} = 1,5$ dvs. H_0 kan inte förkastas på 5 %-nivån. Det finns inget stöd för att sympatierna har ökat efter valet.

26.8

Låt X vara vikten flingor i ett paket. Antag att $X \sim N(\mu, \sigma^2)$ och att observationer på X är stokastiskt oberoende. Vi vill pröva

$H_0 : \mu = 800$

$H_A : \mu < 800$

på signifikansnivån $\alpha = 0,05$. Vi gör $n = 5$ observationer och beräknar urvalsmedelvärdet $\overline{X}$.

Testvariabel: $t = \frac{\overline{X}-800}{S/\sqrt{n}}$

Om H_0 är sann är $t \sim t\,(4)$ (eftersom $n-1=4$).

H_0 förkastas om $t_{obs} < -2,132$

Observerat medelvärde i urvalet är $\overline{x} = \sum x_i/n = 3972/5 = 794,4$

$s^2 = \left(\sum x_i^2 - \left(\sum x_i\right)^2/n\right)/\,(n-1) = (3155754 - 3972^2/5)\,/\,(5-1) = 99,3$

$t_{obs} = \frac{794,4-800}{\sqrt{99,3}/\sqrt{5}} = -1,26$, dvs. H_0 kan inte förkastas på 5 %-nivån.

26.9

Låt X vara antalet korrekt identifierade risker. Då är $X \sim bin\,(n=12,\pi)$. Man testar hypotesen

$H_0 : \pi = 0,5$

$H_A : \pi < 0,5$

och förkastar H_0 om observerat värde $x \leq 3$.

a) $\alpha = P\,(X \leq 3; H_0 \text{ är sann}) = P\,(X \leq 3; \pi = 0,5) = 0,0730$

b) $\beta = P\,(X > 3; \pi = 0,3) = 1 - P\,(X \leq 3; \pi = 0,3) = 1 - 0,4925$. Testets styrka är $1-\beta = 0,4925$.

26.10

Låt P vara urvalsproportionen som föredrar traditionella kassor. I stora urval (stort n) är $P \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$, approximativt, där π är populationsproportionen som föredrar traditionella kassor.

a)

$H_0 : \pi = 0,50$

$H_A : \pi > 0,50$

Signifikansnivå $\alpha = 0,10$

Testvariabel $Z = \frac{P-0,5}{\sqrt{\frac{0,5(1-0,5)}{802}}} = \frac{P-0,5}{0,0177}$

Om H_0 är sann är Z approximativt $N\,(0,1)$ (eftersom $n \geq 30$)

H_0 förkastas om $z_{obs} > 1,28$

Urvalsproportionen observerades till $p = 424/802 = 0,5287$

$z_{obs} = \frac{0,5287-0,50}{0,0177} = 1,63$ dvs. H_0 förkastas på 10 %-nivån. Data stöder hypotesen att fler än 50 % av kunderna föredrar traditionella kassor.

b)

Antag att $\pi = 0,55$. Då är $\frac{\pi(1-\pi)}{n} = \frac{0,55(1-0,55)}{802} = 0,0176^2$ så att $P \sim N(0,55; 0,0176^2)$

$P(\text{Typ II-fel}) = P(t \leq 1,28; \pi = 0,55) = P\left(\frac{P-0,5}{0,0177} \leq 1,28\right) = P(P \leq 0,5221) = P\left(\frac{P-0,55}{0,0126} \leq \frac{0,5221-0,55}{0,0176}\right) \approx P(Z \leq -1,58) \approx 0,06.$

26.11

Låt X vara mängden av den sktiva substansen i en tablett och sätt $E(X) = \mu$.

a) Vi vill pröva

$H_0 : \mu = 500$

$H_A : \mu \neq 500$

Signifikansnivå $\alpha = 0,01$

Testvariabel $t = \frac{\overline{x}-500}{s/\sqrt{n}}$

Antal observationer $n = 250$

Om H_0 är sann är $t \sim N(0,1)$ (approximativt, eftersom $n \geq 30$)

H_0 förkastas om $|t_{obs}| > 2,58$

Vi får $t_{obs} = \frac{499,9-500}{\sqrt{1,2}/\sqrt{250}} \approx -1,44$, dvs. H_0 kan inte förkastas på 1 % nivån.

b) p-värdet är $P(|t| > 1,44) \approx 0,15$. Om H_0 är sann kommer det observerade $|t|$-värdet att bli större än eller lika med det nu observerade $|t|$-värdet i c:a 15 fall av 100.

26.12

Låt X vara relativa förändringar i börskursen. Då är $X \sim N\left(\mu = 0; \sigma^2\right)$.

a) Vi vill pröva

$H_0 : \mu = 0$

$H_A : \mu \neq 0$

Testet blir därmed tvåsidigt.

b) Att förkasta H_0 innebär att teorin att relativa förändringar i börskursen i medeltal är noll inte får stöd av data.

c) $\sigma = 0,01$, $\alpha = 0,05$, $n = 100$. Testvariabel $Z = \frac{\overline{X}-0}{\sigma/\sqrt{n}} = \frac{\overline{X}}{0{,}001} = 1000\overline{X}$.

Förkasta H_0 om $|z_{obs}| > 1,96$, dvs. kritiska värden för Z är -1,96 och 1,96. Kritiska värden för $\overline{X}$ blir därför -0,00196 och 0,00196.

d) Om $\mu = 0,003$ så är $\overline{X} \sim N\left(\mu = 0,003; \sigma = 0,01\right)$ så att $\beta = P\left(-0,00196 < \overline{X} < 0,001\right.$
$P\left(\frac{-0{,}00196-0{,}003}{0{,}01/\sqrt{100}} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{0{,}00196-0{,}003}{0{,}01/\sqrt{100}}\right) = P\left(-4,96 < Z < -1,04\right) \approx 0,15$.

26.13

Låt X vara betalningsviljan hos en person och $E\left(X\right) = \mu$ och $\overline{X}$ är medelbetalningsviljan hos $n = 500$ oberoende personer.

a)

$H_0 : \mu = 120$

$H_A : \mu > 120$

b)

Oberoende observationer. Urvalet är stort $(n > 30)$ så vi kan approximera testvariabelns samplingfördelning med en normalfördelning.

c)

Signifikansnivå $\alpha = 0,05$

Testvariabel $t = \frac{\overline{X}-120}{s/\sqrt{500}}$

Om H_0 är sann är t approximativt $N\left(0,1\right)$ (eftersom $n \geq 30$)

H_0 förkastas om $|t_{obs}| > 1,96$

$\sum x_i = 61000$

$\sum x_i^2 = 8692000$

$\overline{x} = \sum x_i / n = 61000/500 = 122$

$s^2 = \left(\sum x_i^2 - \left(\sum x_i\right)^2 / n\right) / (n-1) = (8692000 - 61000^2/500) / (500-1) = 2505$

$t_{obs} = \frac{122-120}{\sqrt{2505}/\sqrt{500}} = 0,89$, dvs. H_0 kan inte förkastas på 5 %-nivån. Det finns inte stöd för att tro medelbetalningsviljan överstiger 120 kr.

26.14

a) Låt X beteckna en observation på en kvot och låt $E(X) = \mu$. Vi vill pröva hypotesen

$H_0 : \mu = 1$

$H_A : \mu \neq 1$

och antar att X är normalfördelad och att observationer på X är stokastiskt oberoende.

b) Signifikansnivå $\alpha = 0,05$

Testvariabel $t = \frac{\overline{x}-1}{S/\sqrt{n}}$

Om H_0 är sann så är $t \sim t(n-1)$. Förkasta H_0 om $|t_{obs}| > 2,262$

$\overline{x} = \frac{14,63}{10} = 1,463$

$s^2 = \frac{1}{10-1}\left(22,4483 - \frac{14,63^2}{10}\right) \approx 0,1161$

$t_{obs} = \frac{1,463-1}{\sqrt{0,1161}/\sqrt{10}} \approx 4,30$, dvs. H_0 förkastas på 5 procents nivån.

26.15

Låt X vara tillverkningskostnaden per liter. Vi vet att $X \sim N\left(\mu, \sigma_X = 4\right)$

Låt $\overline{X}$ vara medelvärdet av 16 förpackningar så att $\overline{X} \sim N\left(\mu, \sigma_{\overline{X}} = \frac{4}{\sqrt{16}} = 1\right)$

$H_0 : \mu = 10$

$H_A : \mu > 10$

Beslutsregel: H_0 förkastas om $\overline{X} > 11,65$

a)

$P\left(\text{Typ I fel}\right) = P\left(\overline{X} > 11,65; H_0 \text{ sann}\right) = P\left(\frac{\overline{X}-\mu}{\sigma_{\overline{X}}} > \frac{11,65-10}{1}\right) = P\left(Z > 1,65\right) \approx 0,05$

b)

Antag $\mu = 13$

$P\left(\text{Typ II fel}\right) = P\left(\overline{X} < 11,65; \mu = 13\right) = P\left(\frac{\overline{X}-\mu}{\sigma_{\overline{X}}} < \frac{11,65-13}{1}\right) = P\left(Z < -1,35\right) \approx 0,085$

26.16

Låt X vara fokuseringstiden hos en slumpmässigt vald kamera. Vi vet att $X \sim N\left(\mu, \sigma_X = 20\right)$.

Man gör $n = 25$ observationer och låt $\overline{X}$ beteckna medelvärdet av fokuseringstiderna.

a) $H_0 : \mu = 50$

$H_A : \mu < 50$

b) Vi vill använda $\overline{X}$ som testvariabel. Om H_0 är sann så är $\overline{X} \sim N\left(\mu = 50, \sigma_{\overline{X}} = \frac{20}{\sqrt{25}} = 4\right)$

H_0 förkastas om $\overline{x} < k$, där k är kritiskt värde. Med signifikansnivån $\alpha = 0,05$ bestäms det kritiska värdet k genom $P\left(\overline{X} < k\right) = 0,05$, dvs. $P\left(\frac{\overline{X}-50}{4} < \frac{k-50}{4}\right) = 0,05$. Om H_0 är sann är $\frac{\overline{X}-50}{4} \sim N\left(0,1\right)$, vilket innebär att $\Phi\left(\frac{k-50}{4}\right) = 0,05$. Eftersom $\Phi\left(-1,64\right) = 0,05$ följer det att $\frac{k-50}{4} = -1,64$, dvs. $k = 43,44$

c) $\beta = P\left(\overline{X} > k; \mu = 40\right) = P\left(\frac{\overline{X}-40}{4} > \frac{43{,}44-40}{4}\right) = P\left(Z > 0,86\right) = 1 - \Phi\left(0,86\right) \approx 1 - 0,8056 = 0,1944.$

26.17

Låt X vara antal friska. Vi vet att $X \sim bin\left(n = 50, \pi\right)$

$H_0 : \pi = 0,5$

$H_A : \pi = 0,6$

Beslutsregel: Förkasta H_0 om 28 eller fler blir friska

a) $P\left(\text{Typ I fel}\right) = P\left(X \geq 28; H_0 \text{ sann}\right)$

$n\pi = 50 \cdot 0,5 = 25 > 5$ och $n\left(1-\pi\right) = 50 \cdot 0,5 = 25 > 5$ så att approximation med normalfördelning är möjlig

$P\left(\text{Typ I fel}\right) = P\left(\frac{X-n\pi}{\sqrt{n\pi(1-\pi)}} \geq \frac{28-0{,}5-50\cdot0{,}5}{\sqrt{50\cdot0{,}5\cdot0{,}5}}\right) = 1 - P\left(Z \leq 0,71\right) \approx 0,2389$

b) $P\left(\text{Typ II fel}\right) = P\left(X < 28; H_A \text{ sann}\right)$

$n\pi = 50 \cdot 0,6 = 30 > 5$ och $n\left(1-\pi\right) = 50 \cdot 0,4 = 20 > 5$ så att approximation med normalfördelning är möjlig

$P\left(\text{Typ II fel}\right) = P\left(\frac{X-n\pi}{\sqrt{n\pi(1-\pi)}} < \frac{28-0{,}5-50\cdot0{,}6}{\sqrt{50\cdot0{,}6\cdot0{,}4}}\right) = P\left(Z \leq -0,72\right) \approx 0,2358$

26.18

Låt P vara proportionen i urvalet som skulle rösta på Socialdemokraterna och att $E\left(P\right) = \pi$ är proportionen i populationen. Vi vill pröva

$H_0 : \pi = 0,3101$

$H_A : \pi < 0,3101$

med signifikansnivån $\alpha = 0,05$.

Om H_0 är sann så är $P \sim N\left(\mu = 0,3101, \sigma_P^2 = \frac{0{,}3101(1-0{,}3101)}{1921} = 0,000111368\right)$.

Testvariabeln blir därför $Z = \frac{P-0{,}31}{\sqrt{0{,}000111368}}$ och H_0 förkastas om $z_{obs} < -1,64$.

$z_{obs} = \frac{0{,}289-0{,}3101}{\sqrt{0{,}000111368}} \approx -2,00$, dvs. H_0 förkastas på 5 procentsnivån och det finns empiriskt stöd för påståendet att väljarsympatierna hade gått ner efter valet.

26.19

Låt X vara längden av ett armeringsjärn i meter. Vi vet att $X \sim N(\mu, \sigma_X = 0,006)$

$H_0 : \mu = 1$

$H_A : \mu < 1$

$n = 9$ armeringsjärn kontrolleras. $\sigma^2_{\overline{X}} = \sigma^2_X / n = 0,006^2/9 = 0,002^2$ så att $\overline{X} \sim N(\mu, \sigma_{\overline{X}} = 0,002)$

Beslutsregel: H_0 förkastas om $\overline{X} < 0,997$

a) $P(\text{Typ I fel}) = P(X < 0,997; H_0 \text{ sann}) = P\left(\frac{X-\mu}{\sigma_{\overline{X}}} < \frac{0,997-1}{0,002}\right) = P(Z \leq -1,5) \approx 0,0668$

b) n armeringsjärn kontrolleras. $\sigma^2_{\overline{X}} = \sigma^2_X/n = 0,006^2/n$ så att $\overline{X} \sim N(\mu, \sigma_{\overline{X}} = 0,006/\sqrt{n})$

$P(\text{Typ I fel}) = P(X < 0,997; H_0 \text{ sann}) = P\left(\frac{X-\mu}{\sigma_{\overline{X}}} < \frac{0,997-1}{0,006/\sqrt{n}}\right) = P(Z < -0,5\sqrt{n}) = 0,05$

Men $P(Z < -1,64) = 0,05$ så att $-0,5\sqrt{n} = -1,64$

$n = \left(\frac{-1,64}{-0,5}\right)^2 = 10,76$, man behöver minst $n = 11$ observationer.

26.20

Låt X vara tjockleken hos ett slumpmässigt valt papper och låt $\overline{X}$ vara medeltjockleken hos $n = 100$ slumpmässigt valda papper. Vi vet att standardavvikelsen hos X är $\sigma = 0,003$. Enligt Centrala gränsvärdessatsen är då $\overline{X} \sim N\left(\mu, \sigma_{\overline{X}} = \frac{0,003}{\sqrt{100}} = 0,0003\right)$.

a) $H_0 : \mu = 0,06$

$H_A : \mu \neq 0,06$

Testet är tvåsidigt.

b) Maskinen justeras om man förkastar H_0.

c) Maskinen justeras inte om man inte förkastar H_0.

d) Om H_0 är sann så är $\overline{X} \sim N\left(\mu = 0,06, \sigma_{\overline{X}} = 0,0003\right)$. H_0 förkastas om $\overline{X} < k_1$ eller om $\overline{X} > k_2$, där k_1 och k_2 är kritiska värden. De kritiska värdena bestäms genom $P\left(k_1 < \overline{X} < k_2\right) = 0,95$, dvs. $P\left(\frac{k_1 - 0,06}{0,0003} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < \frac{k_2 - 0,06}{0,0003}\right) = 0,95$. Detta är ekvivalent med att $\Phi\left(\frac{k_1 - 0,06}{0,0003}\right) = 0,025$ och $\Phi\left(\frac{k_2 - 0,06}{0,0003}\right) = 0,925$. Eftersom $\Phi(-1,96) = 0,05$ och $\Phi(1,96) = 0,975$ följer det att $\frac{k_1 - 0,06}{0,0003} = -1,96$, dvs. $k_1 = 0,06 - 0,0003 \cdot 1,96 = 0,059412$ och $\frac{k_2 - 0,06}{0,0003} = 1,96$, dvs. $k_2 = 0,06 + 0,0003 \cdot 1,96 = 0,060588$.

e) (i) $\beta = P\left(k_1 < \overline{X} < k_2; \mu = 0,061\right) = P\left(\frac{0,059412 - 0,061}{0,0003} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < \frac{0,060588 - 0,061}{0,0003}\right) \approx P(-5,29 < Z < -1,37) \approx 0,0853$.

(ii) $\beta = P\left(k_1 < \overline{X} < k_2; \mu = 0,0597\right) = P\left(\frac{0,059412 - 0,0597}{0,0003} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < \frac{0,060588 - 0,0597}{0,0003}\right) \approx P(-0,96 < Z < 2,96) \approx 0,9985 - 0,1685 = 0,83$.

27. PRÖVNING AV HYPOTESER OM FLERA POPULATIONSMEDELVÄRDEN

27.1

Låt X_1 och X_2 vara bromssträcka hos respektive biltyp och μ_1 respektive μ_2 medelbromssträcka.

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Signifikansnivå: $\alpha = 5\%$

Testvariabel: $t = \frac{\overline{X}_1 - \overline{X}_2}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$; $S_P^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ är polad varians

Stora urval så att samplingfördelningen för t approximeras med $N(0,1)$

Beslutsregel: H_0 förkastas om $|t_{obs}| > 1,96$

$S_P^2 = \frac{(64-1)6{,}37+(64-1)5{,}44}{64+64-2} = 5,905$; $S_P = \sqrt{5,905} = 2,43$

$t_{obs} = \frac{29{,}50-27{,}25}{2{,}43\sqrt{\frac{1}{64}+\frac{1}{64}}} = 5,24$, dvs. H_0 förkastas på 5 %-nivån. Det finns empiriskt stöd för att det är skillnad mellan biltypernas bromssträcka.

27.2

a) Låt X_1 och X_2 vara lönen hos en slumpmässigt vald man respektive kvinna och låt μ_1 respektive μ_2 vara motsvarande medellöner.

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Testvariabel: $t = \frac{\overline{X}_1 - \overline{X}_2}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$; $S_P^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ är poolad varians

Samplingfördelningen för t är $t\,(n_1+n_2-2)$

Beslutsregel: H_0 förkastas om $|t_{obs}| > t_{krit}$ där t_{krit} är en percentil i t-fördelningen och beror av signifikansnivån.

b) Lönerna antas vara normalfördelade med samma varians för både män och kvinnor. Observationerna antas vara stokastiskt oberoende.

c) Signifikansnivå: $\alpha = 5\%$

$\overline{x}_1 = \frac{257}{9} = 28,56$, $\overline{x}_2 = \frac{272,2}{9} = 30,24$

$s_1^2 = \frac{1}{9-1}\left(7461,96 - \frac{257^2}{9}\right) \approx 15,3978$, $s_2^2 = \frac{1}{9-1}\left(8295,04 - \frac{272,2^2}{9}\right) \approx 7,8128$

$S_P^2 = \frac{(9-1)6,37+(9-1)5,44}{9+9-2} = 11,6053$; $S_P = \sqrt{11,6053} \approx 3,406\,7$

$t_{obs} = \frac{28,56-30,24}{3,4067\sqrt{\frac{1}{9}+\frac{1}{9}}} \approx -1,046\,1$, dvs. H_0 kan inte förkastas på 5 %-nivån. Det finns inget empiriskt stöd för att det är skillnad i medellön för män och kvinnor vid den aktuella produktionsavdelningen.

27.3

Låt Y_1 vara reaktionstiden hos vana Fast Move spelare och Y_2 reaktionstiden hos personer som inte spelar Fast Move. Låt μ_1 och μ_2 vara respektive medelreaktionstid.

a) Observationerna är oberoende av varandra. Antalet observationer i båda urvalen är stora så att approximation med normalfördelningen kan göras.

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 < \mu_2$

b) Signifikansnivå: $\alpha = 5\%$

Testvariabel: $t = \frac{\overline{Y}_1 - \overline{Y}_2}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$; $S_P^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ är polad varians

Stora urval så att samplingfördelningen för t approximeras med $N(0,1)$

Beslutsregel: H_0 förkastas om $t_{obs} < -1,64$

$S_P^2 = \frac{(50-1)0,19+(50-1)0,15}{50+50-2} = 0,17$; $S_P = \sqrt{0,17} = 0,4123$

$t_{obs} = \frac{2,5-2,7}{0,4123\sqrt{\frac{1}{50}+\frac{1}{50}}} = -2,43$, dvs. H_0 förkastas på 5 %-nivån. Det finns empiriskt stöd för att vana Fast Move spelare har kortare reaktionstid.

27.4

Låt X_1 vara tiden det tar att förpacka ett slumpmässigt valt par skor för en anställd med den gamla utbildningen och X_2 vara motsvarande tid för en anställd med den nya utbildningen. Låt μ_1 och μ_2 vara motsvarande medeltider.

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Signifikansnivå: $\alpha = 5\%$

Testvariabel: $t = \frac{\overline{X}_1 - \overline{X}_2}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$; $S_P^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ är polad varians

Samplingfördelningen för t är $t\,(16)$

Beslutsregel: H_0 förkastas om $|t_{obs}| > 2,12$

$\overline{x}_1 = \frac{31{,}7}{9} \approx 3,52$, $\overline{x}_2 = \frac{28{,}4}{9} \approx 3,16$

$s_1^2 = \frac{1}{9-1}\left(113,61 - \frac{31{,}7^2}{9}\right) \approx 0,2444$, $s_2^2 = \frac{1}{9-1}\left(91,22 - \frac{28{,}4^2}{9}\right) \approx 0,2003$

$S_P^2 = \frac{(9-1)6{,}37+(9-1)5{,}44}{9+9-2} \approx 0,2205$; $S_P \approx \sqrt{0,2205} \approx 0,4695$

$t_{obs} = \frac{3.52-3.16}{0.4695\sqrt{\frac{1}{9}+\frac{1}{9}}} \approx 1.63$, dvs. H_0 kan inte förkastas på 5 %-nivån. Det finns inget empiriskt stöd för att det är skillnad i medeltid för de två programmen.

27.5

Vi vill pröva $H_0 : \mu_1 = \mu_2$ mot $H_A : \mu_1 > \mu_2$, där μ_1 och μ_2 är medelavkastningen per delområde. Signifikansnivå $\alpha = 0,05$

Två orelaterade urval med $n_1 = n_2 = 25$ observationer och $\overline{x}_1 = 16,5$, $s_1 = 1,5$, $\overline{x}_2 = 15$ och $s_2 = 1,6$. Poolad varians blir då $s_P^2 = \frac{(25-1)1{,}5^2+(25-1)1{,}6^2}{25+25-2} = 2.405$.

Testvariabel. $t = \frac{\overline{x}_1 - \overline{x}_2}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

Om H_0 är sann så är $t \sim t\,(n_1 + n_2 - 2)$

Beslutsregel: Förkasta H_0 om $t_{obs} > t_{0,95}\,(48) \approx 1,68$

$t_{obs} = \frac{16{,}5-15}{\sqrt{2{,}405}\sqrt{\frac{1}{25}+\frac{1}{25}}} \approx 3,42$, dvs. H_0 förkastas och det kan antas att det nya gödselmedlet är bättre än det gamla.

27.6

Vi vill pröva $H_0 : \mu_1 = \mu_2$ mot $H_A : \mu_1 > \mu_2$, där μ_1 är medelförsäljningen på lördagar och μ_2 är medelförsäljningen på måndagar. Signifikansnivå $\alpha = 0,05$

Två orelaterade urval med $n_1 = n_2 = 25$ observationer och $\overline{x}_1 = 19124$, $s_1 = 5064$, $\overline{x}_2 = 17766$ och $s_2 = 3758$. Poolad varians blir då $s_P^2 = \frac{(25-1)5064^2+(25-1)3758^2}{25+25-2} \approx 4459^2$

Testvariabel. $t = \frac{\overline{x}_1-\overline{x}_2}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

Om H_0 är sann så är $t \sim t\left(n_1 + n_2 - 2\right)$

Beslutsregel: Förkasta H_0 om $t_{obs} > t_{0{,}95}\left(48\right) \approx 1,68$

$t_{obs} = \frac{19124-17766}{4459\sqrt{\frac{1}{25}+\frac{1}{25}}} = 1.08$, dvs. H_0 kan inte förkastas och det finns inga tydliga belägg för att försäljningen är större på lördagar än på måndagar.

27.7

Låt D vara ökning i puls och μ_D vara medelpulsökning.

$H_0 : \mu_D = 0$

$H_A : \mu_D > 0$

Förutsättningar: oberoende observationer mellan individer. Observationerna är normalfördelade.

Signifikansnivå: $\alpha = 5\%$

Antal par av observationer $n = 10$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Om H_0 är sann så är $t \sim t\left(n - 1\right)$

Beslutsregel: H_0 förkastas om $t_{obs} > t_{0{,}95}\left(9\right) = 1,833$

$\overline{d} = \left(24 + 10 + 19 + 25 + 15 + 4 + 12 + 30 + 23 + 30\right)/10 = 19,2$

$s_D^2 = (4376 - 192^2/10) \,/\, (10-1) = 689,6/9 = 76,62$

$t_{obs} = \frac{19{,}2}{\sqrt{76{,}62}/\sqrt{10}} = 6,94$, dvs. H_0 förkastas på 5 %-nivån. Det finns empiriskt stöd för att snus ökar hjärtfrekvensen.

27.8

Låt D vara hastighetsförändringen och anta att D är normalfördelad med väntevärdet μ_D och variansen σ^2. Observationer på D är

22, 4, 11, 10, 1, 8, 37, 9, 23, 1, 9, -1

Antag att observationerna är oberoende av varandra. Vi vill pröva hypotesen

$H_0 : \mu_D = 0$

$H_A : \mu_D > 0$

Signifikansnivå $\alpha = 5\%$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Om H_0 är sann så är $t \sim t\,(11)$.

Beslutsregel: H_0 förkastas om $t_{obs} > 1,796$

$\overline{d} = 134/12 \approx 11,1667$

$s_D^2 = (2848 - 134^2/12) \,/\, (12-1) \approx 122,8788$

$t_{obs} = \frac{11{,}1667}{\sqrt{122{,}8788}/\sqrt{12}} \approx 3,49$, dvs. H_0 förkastas på 5 %-nivån. Det finns empiriskt stöd för att åsynen av polisbilen får bilisterna att sänka farten.

27.9

Låt D vara skillnaden i tentamenspoäng mellan kursdeltagare och ej kursdeltagare och låt μ_D vara medelskillnad i tentamenspoäng.

$H_0 : \mu_D = 0$

$H_A : \mu_D \neq 0$

Förutsättningar: oberoende observationer mellan individer. Observationerna är normalfördelade.

Signifikansnivå: $\alpha = 5\%$

Antal par av observationer $n = 6$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Om H_0 är sann så är $t \sim t\,(n-1)$

Beslutsregel: H_0 förkastas om $|t_{obs}| > t_{0,975}\,(5) = 2,571$

$\overline{d} = (7+2+7+2-1+10)\,/6 = 4,5$

$s_D^2 = \left(207 - 27^2/6\right)/\left(6-1\right) = 85,5/5 = 17,1$

$t_{obs} = \frac{4,5}{\sqrt{17,1}/\sqrt{6}} = 2,67$, dvs. H_0 förkastas på 5 %-nivån. Det finns empiriskt stöd för att den förberedande kursen har effekt på tentamensresultatet.

27.10

Låt D vara förändringen i reaktionstid och anta att D är normalfördelad med väntevärdet μ_D och variansen σ^2. Observationer på D är

-0,07 0,66 0,77 -0,11 0,49 0,70 0,97 0,60 0,57

Antag att observationerna är oberoende av varandra. Vi vill pröva hypotesen

$H_0 : \mu_D = 0$

$H_A : \mu_D \neq 0$

Signifikansnivå $\alpha = 2\%$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Om H_0 är sann så är $t \sim t\,(8)$.

Beslutsregel: H_0 förkastas om $|t_{obs}| > 2,896$

$\overline{d} = 4,58/9 \approx 0,5089$

$s_D^2 = (3,4014 - 4,58^2/9) \, / \, (9-1) \approx 0,1338$

$t_{obs} = \frac{0,5089}{\sqrt{0,1338}/\sqrt{9}} \approx 4,17$, dvs. H_0 förkastas på 2 %-nivån. Det finns empiriskt stöd för att reaktionstiden förändras.

27.11

Låt D vara löneförändring och låt μ_D vara medellöneförändring.

a)Vi vill pröva om kursen innebär att lönen ökar efter genomgången kurs.

$H_0 : \mu_D = 0$

$H_A : \mu_D > 0$

b)Förutsättningar: oberoende observationer mellan individer. Observationerna är normalfördelade.

c)Signifikansnivå: $\alpha = 5\%$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Antal par av observationer $n = 9$

Om H_0 är sann så är $t \sim t\,(n-1)$

Beslutsregel: H_0 förkastas om $t_{obs} > t_{0,95}\,(8) = 1,86$

$\overline{d} = (0,1 - 0,1 + 0 + 0,1 + 0,2 + 0 + 0,2 + 0,3 - 0,1)\,/9 = 0,0778$

$s_D^2 = (0,21 - 0,7^2/9) \, / \, (9-1) = 0,1556/8 = 0,0194$

$t_{obs} = \frac{0,0778}{\sqrt{0,0194}/\sqrt{9}} \approx 1,68$, dvs. H_0 förkastas inte på 5 %-nivån. Det finns inget empiriskt stöd för att kursen påverkar lönen positivt.

27.12

Låt D vara skillnaden i pris mellan Tores livs och Sveahallen och anta att D är normalfördelad med väntevärdet μ_D och variansen σ^2. Observationer på D är

0,45 0,10 0,15 0,20 0,90 0,90 0,15 -0,80 0,10

Antag att observationerna är oberoende av varandra. Vi vill pröva hypotesen

$H_0 : \mu_D = 0$

$H_A : \mu_D \neq 0$

Signifikansnivå $\alpha = 5\%$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Om H_0 är sann så är $t \sim t\,(8)$.

Beslutsregel: H_0 förkastas om $|t_{obs}| > 2,306$

$\overline{d} = 2,15/9 \approx 0,2389$

$s_D^2 = (2,5675 - 2,15^2/9)\,/\,(9-1) \approx 0,2567$

$t_{obs} = \frac{0{,}2389}{\sqrt{0{,}2567}/\sqrt{9}} \approx 1,41$, dvs. H_0 förkastas inte på 5 %-nivån. Det finns inget empiriskt stöd för att priserna är olika.

27.13

Låt D vara skillnaden i handläggningstid mellan avdelningarna och låt μ_D vara medelskillnad i handläggningstid. Vi vill pröva om det finns någon skillnad i handläggningstid mellan avdelningarna.

$H_0 : \mu_D = 0$

$H_A : \mu_D \neq 0$

Förutsättningar: oberoende observationer mellan individer. Observationerna är normalfördelade. Signifikansnivå: $\alpha = 5\%$

Antal par av observationer $n = 8$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Om H_0 är sann så är $t \sim t\,(n-1)$

Beslutsregel: H_0 förkastas om $|t_{obs}| > t_{0{,}975}\,(7) = 2,365$

$\overline{d} = (1 - 2 + 0 - 2 + 1 - 2 + 1 + 0)\,/8 = -0,375$

$s_D^2 = \left(15 - (-3)^2/8\right)/\,(8-1) = 13,875/7 = 1,9821$

$t_{obs} = \frac{-0{,}375}{\sqrt{1{,}9821}/\sqrt{8}} = -0,7534$, dvs. H_0 förkastas inte på 5 %-nivån. Det finns inget empiriskt stöd för att det föreligger någon skillnad mellan avdelningarna.

27.14

Låt D vara skillnaden mellan TBs och SWs värdering av en lägenhet och anta att D är normalfördelad med väntevärdet μ_D och variansen σ^2. Observationer på D är

45 10 15 20 90 90 15 -80 10 (tusentals kr)

Antag att observationerna är oberoende av varandra. Vi vill pröva hypotesen

$H_0 : \mu_D = 0$

$H_A : \mu_D \neq 0$

Signifikansnivå $\alpha = 5\%$

Testvariabel: $t = \frac{\overline{D}}{S_D/\sqrt{n}}$

Om H_0 är sann så är $t \sim t\,(8)$.

Beslutsregel: H_0 förkastas om $|t_{obs}| > 2,306$

$\overline{d} = 215/9 \approx 23,8889$

$s_D^2 = \left(25675 - 215^2/9\right) / \left(9 - 1\right) \approx 2567,361$

$t_{obs} = \frac{23{,}8889}{\sqrt{2567{,}361}/\sqrt{9}} \approx 1,41$, dvs. H_0 förkastas inte på 5 %-nivån. Det finns inget empiriskt stöd för att TB och SW värderar lägenheterna olika.

27.15

Vi vill pröva $H_0 : \mu_1 = \mu_2 = \mu_3$ mot att minst ett μ_j skiljer sig från de andra, där μ_j är förväntad poäng på tentamen i grupp (behandling) j. Signifikansnivå $\alpha = 0,05$

Vi har $n_1 = n_2 = n_3 = 10$ observationer och medelvärdena i grupperna är $\overline{x}_1 = 62,7$, $\overline{x}_2 = 63,3$ och $\overline{x}_3 = 61,7$. Medelvärdet av alla observationer är $\overline{x}_{.} = 62,5667$.

$SS_A = (60 - 62,7)^2 + (61 - 62,7)^2 + \ldots + (69 - 61,7)^2 = 1846,3$, $SS_0 = (60 - 62,5667)^2 + (61 - 62,5667)^2 + \ldots + (69 - 62,5667)^2 = 1859,367$

Testvariabel. $F = \frac{(SS_0 - SS_A)/(3-1)}{SS_A/(30-3)}$

Om H_0 är sann så är $F \sim F\,(3-1, n_1+n_2+n_3-3)$

Beslutsregel: Förkasta H_0 om $F_{obs} > F_{0,95}\,(2,27) = 3,35$

Variation	SS	fg	MS	F
Mellan grupper	13,067	2	6,533	0,096
Inom grupper	1846,300	27	68,381	
Totalt	1859,367	29		

$F_{obs} = \frac{(1859{,}367-1846{,}3)/2}{1846{,}3/27} = 0,096$ dvs. H_0 kan inte förkastas. Det finns inet empiriskt evidens för att grupperna skiljer sig åt i tentamenspoäng

27.16

Vi vill pröva $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ mot att minst ett μ_j skiljer sig från de andra, där μ_j är förväntad bensinförbrukning för biltyp j. Signifikansnivå $\alpha = 0,05$

Vi antar att bensinförbrukningen är normalfördelad med samma varians för alla biltyper och att observationerna är stokastiskt oberoende.

Vi har $n_1 = n_2 = n_3 = n_4 = 10$ observationer och medelvärdena i grupperna är $\overline{x}_1 = 0,70$, $\overline{x}_2 = 0,64$ och $\overline{x}_3 = 0,67$ och $\overline{x}_4 = 0,65$. Eftersom alla grupper har samma antal observationer blir medelvärdet av alla observationer $\overline{x}_. = (0,70+0,64+0,67+0,65)/4 = 0,665$. Varianserna i grupperna observerades till $s_1^2 = 0,0025$, $s_2^2 = 0,0019$, $s_3^2 = 0,0024$ och $s_4^2 = 0,0022$.

Detta ger $SS_A = (16-1)\,0.0025 + (16-1)\,0.0019 + (16-1)\,0.0024 + (16-1)\,0.0022 = 0,135$ och $SS_0 - SS_A = 16\,(0.70-0.665)^2 + 16\,(0.64-0.665)^2 + 16\,(0.67-0.665)^2 + 16\,(0.65-0.665)^2 = 0.0336$.

Testvariabel. $F = \frac{(SS_0 - SS_A)/(4-1)}{SS_A/(64-4)}$

Om H_0 är sann så är $F \sim F\,(4-1, n_1+n_2+n_3+n_4-4)$

Beslutsregel: Förkasta H_0 om $F_{obs} > F_{0,95}\,(3,60) = 2,76$

Variation	SS	fg	MS	F
Mellan grupper	0,0336	3	0,0112	4,98
Inom grupper	0,1350	60	0,00225	
Totalt	0,1686	63		

$F_{obs} = \frac{0,0336/3}{0,1350/60} = 4,98$ dvs. H_0 förkastas på 5% nivån. Det finns ett empiriskt evidens för att bensinförbrukningen är olika för de olika biltyperna.

27.17

Vi vill pröva $H_0 : \mu_1 = \mu_2 = \mu_3$ mot att minst ett μ_j skiljer sig från de andra, där μ_j är förväntad värdering som mäklare j gör. Signifikansnivå $\alpha = 0,05$

Vi antar att värderingarna är normalfördelad med samma varians för alla mäklare och objekt och att observationerna är stokastiskt oberoende.

Observationerna på ett objekt är relaterade, så vi har $n = 9$ block och $m = 3$ mäklare.

Med observationer uttryckta i miljoner kronor får vi $SS_A = 0,025311$ och $SS_0 = 0,043383$ så att $SS_0 - SS_A = 0,043383 - 0,025311 = 0,018072$.

Testvariabel. $F = \frac{(SS_0 - SS_A)/(3-1)}{SS_A/(9 \cdot 3 - 9 - 3 + 1)}$

Om H_0 är sann så är $F \sim F\,(2,16)$

Beslutsregel: Förkasta H_0 om $F_{obs} > F_{0,95}\,(2,16) = 3,63$

$F_{obs} = \frac{0,018072/2}{0,025311/16} = 5,71$ dvs. H_0 förkastas. Det finns empiriskt stöd för att mäklarna gör olika värderingar av sammaobjekt.

27.18

Vi vill pröva $H_0 : \mu_1 = \mu_2 = \mu_3$ mot att minst ett μ_j skiljer sig från de andra, där μ_j är förväntad handläggningstid på avdelning j. Signifikansnivå $\alpha = 0,05$

Vi antar att värderingarna är normalfördelad med samma varians för alla observationer att observationerna är stokastiskt oberoende.

Observationerna på ett objekt är relaterade, så vi har $n = 8$ ärenden (block) och $m = 3$ avdelningar.

Vi får $SS_A = 11,417$ och $SS_0 = 18$ så att $SS_0 - SS_A = 18 - 11,417 = 6,583$.

Testvariabel. $F = \frac{(SS_0 - SS_A)/(3-1)}{SS_A/(8 \cdot 3 - 8 - 3 + 1)}$

Om H_0 är sann så är $F \sim F\,(2, 14)$

Beslutsregel: Förkasta H_0 om $F_{obs} > F_{0{,}95}\,(2, 14) = 3{,}74$

$F_{obs} = \frac{6.583/2}{11.417/14} = 4.036$ dvs. H_0 förkastas på 5% nivån. Det finns empiriskt stöd för att handläggningstiden är olika på de tre avdelningarna.

28. LINJÄR REGRESSION MED EN FÖRKLARINGSVARIABEL

28.1

Låt x vara kroppsvikt utan fett och Y energiomsättning. Anta att $Y \sim N(\mu_x, \sigma^2)$, där $\mu_x = \alpha + \beta x$ och att observationer på x och Y är stokastiskt oberoende. Från exempel 19.2 vet vi att $\overline{x} = \frac{516,5}{12} = 43,042$, $\overline{Y} = \frac{14881}{12} = 1240,08$, $s_{xY} = 1127,5780$, $s_x^2 = 46,8936$. Vi får då parameteruppskattningarna av β och α som $b = \frac{s_{xY}}{s_x^2} = \frac{1127,5780}{46.8936} = 24,045$ respektive $a = \overline{Y} - b\overline{x} = 1240,08 - 24,045 \cdot 43,042 = 205,13$. Residualerna kan nu beräknas genom $e_i = Y_i - (a + bx)$ vilket ger $SS_E = \sum e_i^2 = 87732,25$. Vidare vet vi att $SS_T = (n-1)\, s_Y^2 = (12-1)\, 35088,8106 = 385976$. så att $SS_R = SS_T - SS_E = 385976 - 87732,25 = 298243,75$ så att förklaringsgraden blir $R^2 = \frac{SS_R}{SS_T} = \frac{298243,75}{385976} = 0,773$. En uppskattning av σ^2 som $s_e^2 = \frac{SS_E}{n-2} = \frac{87732,25}{12-2} = 8773,2$.

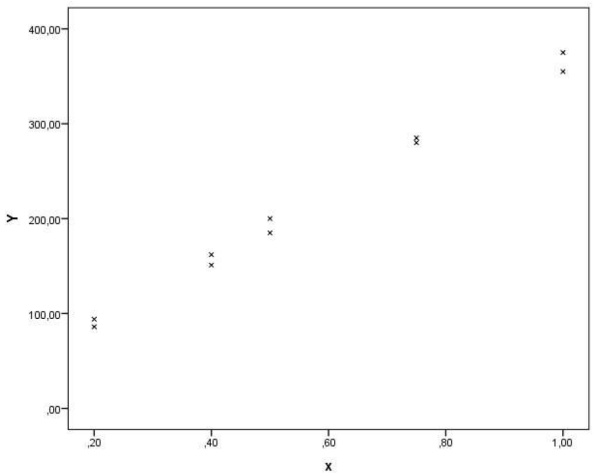
Vi testar nu $H_0 : \beta = 0$. Testvariabel är $t = \frac{b}{s_b}$ och H_0 förkastas på 5 % nivån om $|t_{obs}| > t_{0,975}\,(n-2) = 2,228$. Det gäller att $s_b^2 = \frac{s_e^2}{(n-1)s_x^2} = \frac{8773,2}{(12-1)46,8936} = 17,008$ så att $t_{obs} = \frac{24,045}{\sqrt{17,008}} = 5,83$, dvs. H_0 förkastas på 5% nivån och det finns empiriskt evidens för att kroppsvikt utan fett linjärt påverkar förväntad energiomsättning. Prediktionsintervall för energiomsättningen fås genom $\widehat{\mu}_x \pm t_{0,975}\,(n-2)\, s_0$ där $s_0^2 = s_e^2 \left(1 + \frac{1}{n} + \frac{(x-\overline{x})^2}{(n-1)s_x^2}\right)$. För $x = 43$ får vi $\widehat{\mu}_{43} = 205,13 + 24,045 \cdot 43 = 1239,1$ och $s_{43}^2 = 8773,2 \left(1 + \frac{1}{12} + \frac{(43-43,042)^2}{(12-1)46,8936}\right) = 9504,3$ så att ett prediktionsintervall för Y då $x = 43$ är $1239,1 \pm 2,228\sqrt{9504,3}$ dvs. $1239,1 \pm 217,21$. För För $x = 50$ får vi $\widehat{\mu}_{50} = 205,13 + 24,045 \cdot 50 = 1407,4$ och $s_{50}^2 = 8773,2 \left(1 + \frac{1}{12} + \frac{(50-43,042)^2}{(12-1)46,8936}\right) = 10328$ så att ett prediktionsintervall för Y då $x = 50$ är $1407,4 \pm 2,228\sqrt{10328}$ dvs. $1407,4 \pm 226,42$.

28.2

Låt x vara gödselmängd och Y skörd. Vi får först att $\sum x = 2800$, $\sum x^2 = 1400000$, $\sum Y = 23520$, $\sum Y^2 = 83260850$ och $\sum xY = 10472500$. Detta ger $b = \frac{10472500 - \frac{2800 \cdot 23520}{7}}{1400000 - \frac{2800^2}{7}} = 3,8018$ och $a = \frac{23520}{7} - 3,8018 \frac{2800}{7} = 1839,3$. Vidare får vi $SS_T = 83260850 - \frac{23520^2}{7} = 4233650$. Residualer beräknas genom $e_i = Y_i - (a + bx)$ vilket ger $SS_E = \sum e_i^2 = 186650$.och följaktligen $SS_R = SS_T - SS_E = 4233650 - 186650 = 4047000 = 4047000$. Detta ger $R^2 = \frac{4047000}{4233650} = 0,956$. För att pröva hypotesen $H_0 : \beta = 0$ beräknar vi först $s_e^2 = \frac{SS_E}{n-2} = \frac{186650}{7-2} = 37330$ och $s_b^2 = \frac{37330}{1400000 - \frac{2800^2}{7}} = 0,13332$ vilket ger testvariabeln $t_{obs} = \frac{3,8018}{\sqrt{0,13332}} = 10,412$. H_0 förkastas på 5% nivån om $|t_{obs}| > t_{0,975}(7-2) = 2,571$. Vi ser att H_0 förkastas på 5% nivån och vi drar slutsatsen att gödselmängden påverkar förväntad mängd skörd. Då $x = 400$ får vi uppskattningan av förväntad skörd till $\widehat{\mu}_{400} = 1839,3 + 3,8018 \cdot 400 = 3360$ med uppskattad varians $37330 \left(\frac{1}{7} + \frac{\left(400 - \frac{2800}{7}\right)^2}{1400000 - \frac{2800^2}{7}} \right) = 5332,9$ så att ett konfidensintervall blir $3360 \pm 2,571\sqrt{5332,9}$, dvs. $3360 \pm 187,75$. Då $x = 400$ får vi $s_0^2 = 37330 \left(1 + \frac{1}{7} + \frac{\left(400 - \frac{2800}{7}\right)^2}{1400000 - \frac{2800^2}{7}} \right) = 42663$ så att ett prediktionsintervall för Y ges av $3360 \pm 2,571\sqrt{42663}$ dvs. 3360 ± 531.

28.3

Låt x vara mängd vatten i liter och y vara tiden till kokning i sekunder. En figur över observationer på x och y är



En modell över sambandet mellan x och y är $y_i = \alpha + \beta x_i + \varepsilon_i$

Vi får att $\sum x_i = 5,7$, $\sum x_i^2 = 4,025$, $\sum y_i = 2173$, $\sum x_i y_i = 1507,45$, $\sum y_i^2 = 565777$.

Minstakvadratskattningar av α och β är

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{1507{,}45 - 5{,}7 \cdot 2173/10}{4{,}025 - 5{,}7^2/10} = 346,44$

$a = \overline{y} - b\overline{x} = \frac{2173}{10} - 346,44 \frac{5{,}7}{10} = 19,829$

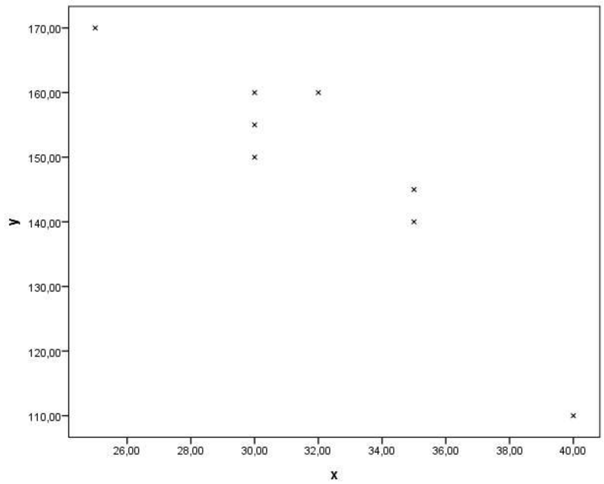
Residualerna beräknas genom $e_i = y_i - (19,829 + 346,44 \cdot x_i)$ Summan av kvadrerade residualer blir $SSE = \sum e_i^2 = 446,28$. Totala kvadratsumman är $SST = \sum (y_i - \overline{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 565777 - \frac{2173^2}{10} = 93584,1$. Vi får därför $SSR = SST - SSE = 93584.1 - 446.28 = 93138$ och $R^2 = \frac{SSR}{SST} = \frac{93138}{93584.1} = 0.995$.

a är en uppskattning av tiden i sekunder det tar att värma ett tomt kärl till 100 grader. b är en uppskattning av hur mycket längre tid i sekunder det tar att värma en liter vatten.

Det finns ett kausalt samband mellan vattenmängd och tid till kokning.

28.4

Låt x vara pris och y vara såld kvantitet. En figur över observationer på x och y är



En modell över sambandet mellan x och y är $y_i = \alpha + \beta x_i + \varepsilon_i$

Vi får att $\sum x_i = 257$, $\sum x_i^2 = 8399$, $\sum y_i = 1190$, $\sum x_i y_i = 37695$, $\sum y_i^2 = 179350$.

Minstakvadratskattningar av α och β är

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{37695 - 257 \cdot 1190/8}{8399 - 257^2/8} = -3,7358$

$a = \overline{y} - b\overline{x} = \frac{1190}{8} - (-3,7358)\,\frac{257}{8} = 268,76$

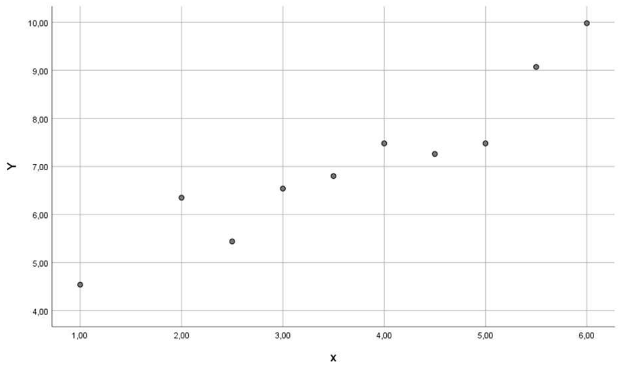
Residualerna kan nu beräknas genom $e_i = y_i - (268,76 - 3,7358 \cdot x_i)$. Summan av kvadrerade residualer blir $SSE = \sum e_i^2 = 343,53$. Totala kvadratsumman är $SST = \sum (y_i - \overline{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 179350 - \frac{1190^2}{8} = 2337.5$. Vi får därför $SSR = SST - SSE = 2337.5 - 343.53 = 1994.0$ och $R^2 = \frac{SSR}{SST} = \frac{1994.0}{2337.5} = 0.853$

a har ingen meningsfull tolkning. b är en uppskattning av hur mycket försäljningen ökar då priset ökar med en krona.

Sambandet mellan x och y är inte kausalt.

28.5

Låt x vara ålder och Y vikten hos små barn. Beteckna väntevärdet för Y med μ_x. Ett linjärt samband mellan färväntad vikt och ålder är då $\mu_x = \alpha + \beta x$. En tolkning av α är födelsevikten och en tolkning av β är viktökning under en nmånad. Vi har $n = 10$ observationer. Ett punktdiagram över observationerna är



Vi får att $\sum x_i = 37$, $\sum x_i^2 = 160$, $\sum Y_i = 70,94$, $\sum Y_i^2 = 526,01$ och $\sum x_i Y_i = 284,02$.

Minstakvadratuppskattning av α och β är

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{284{,}02 - 37 \cdot 70{,}94/10}{160 - 37^2/10} = 0,93255$

$a = \overline{y} - b\overline{x} = \frac{70{,}94}{10} - 0,93255 \frac{37}{10} = 3,6436$

För att pröva hypotesen $H_0 : \beta = 0$ antar vi att observationerna är normalfördelade och stokastiskt oberoende. Vi beräknar först residualer $e_i = Y_i - 3,6436 - 0,93255 x_i$ och en uppskattning av residualvariansen $s_e^2 = \sum e_i^2 / (n-2) = 0,336$. Detta ger en uppskattning av variansen av b som $s_b^2 = \frac{s_e^2}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{0{,}336}{160 - \frac{37^2}{10}} = 0,014545$.

Testvariabeln beräknas till $t_{obs} = \frac{b}{s_b} = \frac{0{,}93255}{\sqrt{0{,}014545}} = 7,732\,4$. H_0 förkastas på 5% nivån om $|t_{obs}| > t_{0{,}975}(10-2) = 2,306$. Vi ser att H_0 förkastas på 5%

nivån, dvs. det finns ett empiriskt evidens att vikten ökar med ålder hos små barn.

Konfidensintervall för μ_x då $x = 1$ ges av $\widehat{\mu}_1 \pm t_{0,975}\,(8)\sqrt{\frac{1}{n} + \frac{(1-\overline{x})^2}{(n-1)s_x^2}}$, vi får intervallet $3,6436 + 0,93255 \cdot 1 \pm 2,306\sqrt{\frac{1}{10} + \frac{(1-3,7)^2}{23,1}}$, dvs. $4,5762 \pm 1,486\,6$.

För $x = 3,5$ får vi på motsvarande sätt $3,6436{+}0,93255{\cdot}3,5{\pm}2,306\sqrt{\frac{1}{10} + \frac{(3,5-3,7)^2}{23,1}}$, dvs. $6,9075 \pm 0,7355$

och för $x = 6$ får vi $3,6436 + 0,93255 \cdot 6 \pm 2,306\sqrt{\frac{1}{10} + \frac{(6-3.7)^2}{23,1}}$, dvs. $9,2389 \pm 1,3227$.

Ett prediktionsintervall för Y då $x = 3,5$ ges av $\widehat{\mu}_1{\pm}t_{0,975}\,(8)\sqrt{1 + \frac{1}{n} + \frac{(3,5-\overline{x})^2}{(n-1)s_x^2}}$.

Detta ger prediktionsintervallet $3,6436{+}0,93255{\cdot}3,5{\pm}2,306\sqrt{1 + \frac{1}{10} + \frac{(3,5-3,7)^2}{23,1}}$, dvs. $6,9075 \pm 2,4205$

28.6

Vi vet att $n = 22$, $a = 5,2$, $b = -4,4$, $SS_E = 140$ och $SS_R = 260$.

Då är $SS_T = SS_E + SS_R = 140 + 260 = 400$ så att $R^2 = \frac{SS_R}{SS_T} = \frac{260}{400} = 0,65$.

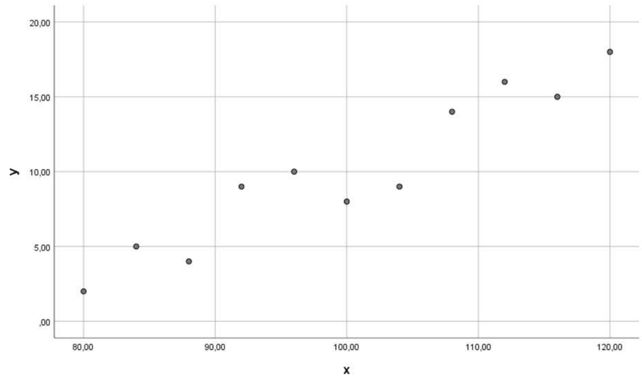
För att pröva hypotesen $H_0 : \beta = 0$ kan vi använd testvariabeln $F = \frac{SS_T - SS_E}{SS_E/(n-2)}$ och förkasta H_0 på 1% nivån om $F_{obs} > F_{0,99}\,(1, n-2) = F_{0,99}\,(1, 20) = 8,10$. Vi får att $F_{obs} = \frac{400-140}{140/(22-2)} = 37,143$, dvs. H_0 förkastas på 1% nivån.

Alternativt kan vi använda testvariabeln $t = \frac{b}{s_b}$, där $s_b^2 = \frac{s_e^2}{(n-1)s_x^2}$, $s_e^2 = \frac{SS_E}{n-2}$. H_0 förkastas på 1% nivån om $|t_{obs}| > t_{0,995}\,(n-2) = t_{0,995}\,(20) = 2,845$.

Vi ser först att $s_e^2 = \frac{140}{22-2} = 7$. Eftersom $b = r_{x,Y}\frac{s_Y}{s_x}$ gäller det att $(n-1)\,s_x^2 = r_{x,Y}^2\frac{(n-1)s_Y^2}{b^2} = R^2\frac{SS_T}{b^2} = 0,65\frac{400}{(-4,4)^2} = 13,430$. Detta ger $s_b^2 = \frac{7}{13,430} = 0,52122$ och $t_{obs} = \frac{-4,4}{\sqrt{0,52122}} = -6,0946$, dvs. H_0 förkastas på 1% nivån.

28.7

Låt x vara temperatur och Y syremängd. Beteckna väntevärdet av Y med μ_x. Ett linjärt samband mellan x och Y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



Det gäller att $\sum x_i = 1100$, $\sum x_i^2 = 111760$, $\sum Y_i = 110$, $\sum Y_i^2 = 1372$ och $\sum x_i Y_i = 11660$ och $n = 11$. Minstakvadratuppskattningar av α och β ges då av

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{11660 - 1100 \cdot 110/11}{111760 - 1100^2/11} = 0.375$

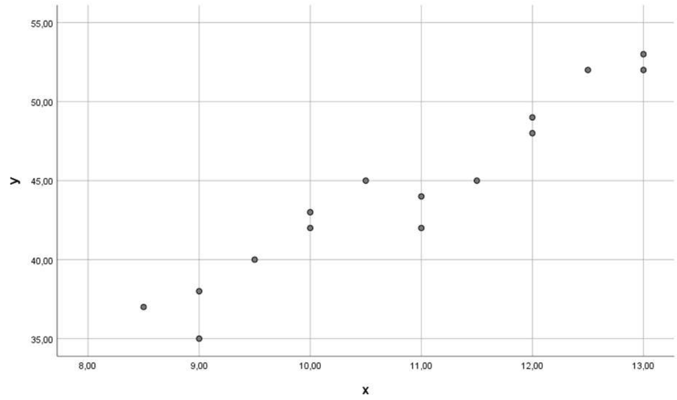
$a = \overline{y} - b\overline{x} = \frac{110}{11} - 0.375 \cdot \frac{1100}{11} = -27,5$

För att pröva hypotesen $H_0 : \beta = 0$ antar vi att observationerna är normalfördelade och stokastiskt oberoende. Vi beräknar först residualer $e_i = Y_i - (-27,5 + 0,375 x_i)$ och en uppskattning av residualvariansen $s_e^2 = \sum e_i^2 / (n-2) = 24,5/(11-2) = 2,722$. Detta ger en uppskattning av variansen av b som $s_b^2 = \frac{s_e^2}{\sum x_i^2 - \left(\sum x_i\right)^2/n} = \frac{2.722}{111760 - \frac{1100^2}{11}} = 0,0015466$.

Testvariabeln beräknas till $t_{obs} = \frac{b}{s_b} = \frac{0.375}{\sqrt{0.0015466}} = 9,5355$. H_0 förkastas på 5% nivån om $|t_{obs}| > t_{0,975}(11-2) = 2,262$. Vi ser att H_0 förkastas på 5% nivån, dvs. det finns ett empiriskt evidens för att temperaturen påverkar mängden syre i slutprodukten.

28.8

Låt x vara löptempo och Y VOmax. Beteckna väntevärdet av Y med μ_x. Ett linjärt samband mellan x och Y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



Det gäller att $\sum x_i = 162,5$, $\sum x_i^2 = 1791,25$, $\sum Y_i = 665$, $\sum Y_i^2 = 29923$ och $\sum x_i Y_i = 7316,5$ och $n = 15$. Minstakvadratuppskattningar av α och β ges då av

$$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{7316.5 - 162.5 \cdot 665/15}{1791.25 - 162.5^2/15} = 3,6432$$

$$a = \overline{y} - b\overline{x} = \frac{665}{15} - 3.6432 \cdot \frac{162.5}{15} = 4,8653$$

För att pröva hypotesen $H_0 : \beta = 0$ antar vi att observationerna är normalfördelade och stokastiskt oberoende. Vi beräknar först residualer $e_i = Y_i - (4,8653 + 3,6432 x_i)$ och en uppskattning av residualvariansen $s_e^2 = \sum e_i^2 / (n-2) = 32,076/(15-2) = 2,467$. Detta ger en uppskattning av variansen av b som $s_b^2 = \frac{s_e^2}{\sum x_i^2 - \left(\sum x_i\right)^2/n} = \frac{2.467}{1791.25 - \frac{162.5^2}{15}} = 0,080011$.

Testvariabeln beräknas till $t_{obs} = \frac{b}{s_b} = \frac{3.6432}{\sqrt{0.080011}} = 12,880$. H_0 förkastas på 5% nivån om $|t_{obs}| > t_{0,975}(15-2) = 2,16$. Vi ser att H_0 förkastas på 5% nivån, dvs. det finns ett empiriskt evidens för att löptempo påverkar VOmax.

Ett 95 %-igt konfidensintervall för β ges av $b \pm t_{0,975}(n-2)\, s_b$, detta ger oss konfidensintervallet $3.6432 \pm 2.16 \cdot \sqrt{0.080011}$, dvs. $3.6432 \pm 0,61098$.

28.9

Låt x vara mängden av födoämnet och y vara viktökning. Beteckna väntevärdet av y med μ_x.Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$.

a) Enligt uppgiften gäller det att $\sum y_i = 150$, $\sum x_i = 180$, $\sum (y_i - \overline{y})^2 = 295$,

$\sum (x_i - \overline{x})^2 = 295$, $r_{x,y} = 0,7$, $s^2 = 2,59$ och $n = 60$. Vi börjar med att uppskatta parametrarna i det linjära sambandet $\mu_x = \alpha + \beta x$. Vi har ingen uppgift om $\sum (x_i - \overline{x})(y_i - \overline{y})$, som är täljaren i uttrycket för b, men vi vet att $r_{x,y} = \sum (x_i - \overline{x})(y_i - \overline{y}) / \sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}$, så att

$$\sum (x_i - \overline{x})(y_i - \overline{y}) = r_{x,y}\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2} = 0,7\sqrt{295 \cdot 295} = 206,5.$$

$b = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{206.5}{295} = 07$

$a = \overline{y} - b\overline{x} = \frac{150}{60} - 0.7 \cdot \frac{180}{60} = 0.4$

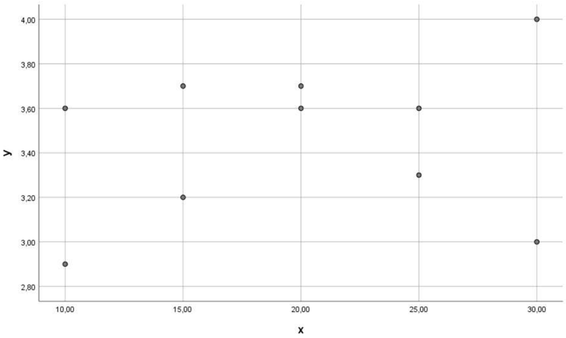
En uppskattning av den förväntade viktökningen om $x = 1,25$ blir därför $0.4 + 0.7 \cdot 1.25 = 1.275$.

b) En uppskattning av variansen för b är $s^2 / \sum (x_i - \overline{x})^2 = 2.59/295 = 0,0087797$. Eftersom n är stort ges ett 95 %-igt konfidensintervall för β av $0.7 \pm 1.96 \cdot \sqrt{0.0087797}$, dvs. $0.7 \pm 0,18365$.

c) En uppskattning av den förväntade viktökningen om $x = 3,0$ är $0.4 + 0.7 \cdot 3.0 = 2.5$. Variansen för en ny observation vid $x = 3,0$ är $2.59\left(1 + \frac{1}{60} + \frac{\left(3.0 - \frac{150}{60}\right)^2}{295}\right) = 2,6354$ så att ett 95 %-igt prediktionsintervall ges av $2,5 \pm 1.96\sqrt{2.6354}$, dvs. $2,5 \pm 3,1818$.

28.10

Låt x vara mängden av temperatur och y vara hur djupt impregneringsvätskan har tränt in. Beteckna väntevärdet av y med μ_x.Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



a) I uppgiften är det givet att $\sum x_i^2 = 4500$, $\sum y_i^2 = 120,8$, $\sum x_i y_i = 697$ och $n = 10$. Från givna data betäknar vi $\sum x_i = 200$ och $\sum y_i = 34,6$. Vi ska uppskatta parametrarna α och β.

$$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{697 - 200 \cdot 34.6/10}{4500 - 200^2/10} = 0.01$$

$$a = \overline{y} - b\overline{x} = \frac{34.6}{10} - 0.01 \cdot \frac{200}{10} = 3.26$$

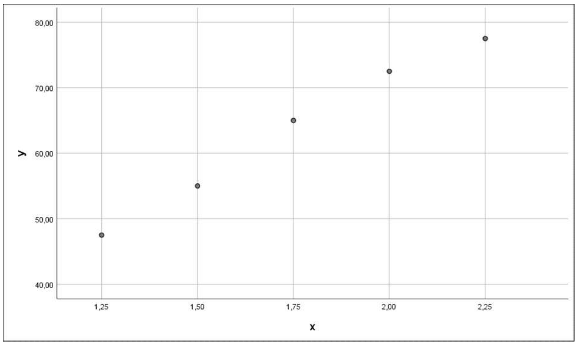
b) För att pröva hypotesen $H_0 : \beta = 0$ antar vi att observationerna på y är normalfördelade och stokastiskt oberoende. Vi beräknar först residualer $e_i = Y_i - (3,26 + 0,01x_i)$ och en uppskattning av residualvariansen $s_e^2 = \sum e_i^2 / (n-2) = SS_E / (n-2) = 1,034/(10-2) = 0,129$. Detta ger en uppskattning av variansen av b som $s_b^2 = \frac{s_e^2}{\sum x_i^2 - \left(\sum x_i\right)^2/n} = \frac{0.129}{4500 - \frac{200^2}{10}}. = 0,000258$

Testvariabeln beräknas till $t_{obs} = \frac{b}{s_b} = \frac{0.01}{\sqrt{0.000258}} = 0,6226$. H_0 förkastas på 5% nivån om $|t_{obs}| > t_{0,975}(10-2) = 2,306$. Vi ser att H_0 kan inte förkastas på 5% nivån, dvs. det finns inget starkt empiriskt evidens för att temperaturen påverkar djupet av impregneringen linjärt.

Alternativt kan vi beräkna $SS_T = \sum y_i^2 - \left(\sum y_i\right)^2 = 120.8 - 34.6^2/10 = 1,084$ och sedan använda testvariabeln $F_{obs} = \frac{1,084 - 1,034}{1.034/(10-2)} = 0,38685$. H_0 förkastas om $F > F_{0,05}(1, 10-2) = 5,32$, dvs. H_0 kan inte förkastas på 5 %-nivån. Observeta att $F_{obs} = t_{obs}^2$.

28.11

a) Låt x vara kiselhalt och y vara hårdhet hos järnet. Beteckna väntevärdet av y med μ_x. Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



b) Vi ska uppskatta parametrarna i det linjära sambandet. Från de givna observationerna får vi $\sum x_i = 8,75$, $\sum x_i^2 = 15,9375$, $\sum y_i = 317,5$, $\sum y_i^2 = 20768,75$, $\sum x_i y_i = 575$ och $n = 5$. Vi får därför

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{575 - 8.75 \cdot 317.5/5}{15.9375 - 8.75^2/5} = 31.0$

$a = \overline{y} - b\overline{x} = \frac{317.5}{5} - 31 \cdot \frac{8.75}{5} = 9.25$

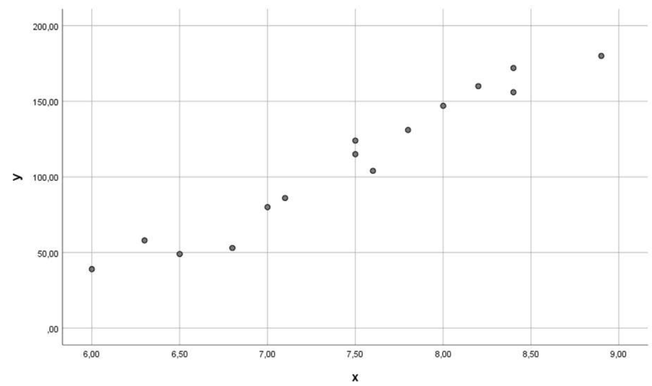
c) För att pröva hypotesen $H_0 : \beta = 0$ antar vi att observationerna på y är normalfördelade och stokastiskt oberoende. Vi beräknar först residualer $e_i = Y_i - (9,25 + 31x_i)$ och en uppskattning av residualvariansen $s_e^2 = \sum e_i^2 / (n-2) = SS_E / (n-2) = 6,875 / (3-2) = 2,294$. Detta ger en uppskattning av variansen av b som $s_b^2 = \frac{s_e^2}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{2.294}{15.9375 - \frac{8.75^2}{5}} = 3.6704$

Testvariabeln beräknas till $t_{obs} = \frac{b}{s_b} = \frac{31.0}{\sqrt{3.6704}} = 16.181$. H_0 förkastas på 5% nivån om $|t_{obs}| > t_{0,975}(5-2) = 3,182$. Vi ser att H_0 förkastas på 5% nivån, dvs. det finns ett empiriskt stöd för att kiselhalten påverkar hårdheten.

Alternativt kan vi beräkna $SS_T = \sum y_i^2 - \left(\sum y_i\right)^2 = 20768.75 - 317.5^2/5 = 607.5$ och sedan använda testvariabeln $F_{obs} = \frac{607.5 - 6.875}{6.875/(5-2)} = 262.09$. H_0 förkastas om $F > F_{0,05}(1, 5-2) = 10,1$, dvs. H_0 kan inte förkastas på 5 %-nivån.

28.12

a) Låt x vara föroreningsgrad och y vara instrumentets utslag. Beteckna väntevärdet av y med μ_x. Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



Vi uppgift är att uppskatta parametrarna i det linjära sambandet. Från de givna observationerna får vi $\sum x_i = 112$, $\sum x_i^2 = 846,26$, $\sum y_i = 1654$, $\sum y_i^2 = 213998$, $\sum x_i y_i = 12898,80$ och $n = 15$ vilket ger uppskattningarna

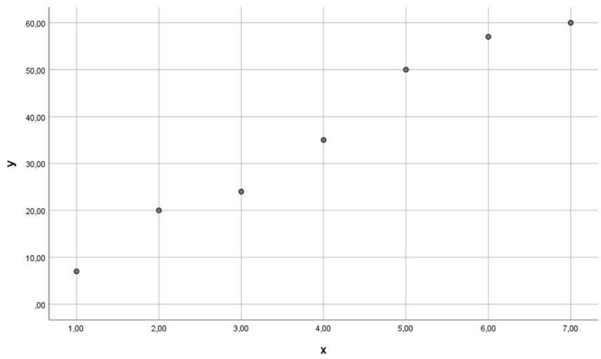
$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{12898.80 - 112 \cdot 1654/15}{846.26 - 112^2/15} = 54,930$

$a = \overline{y} - b\overline{x} = \frac{1654}{15} - 54.93 \cdot \frac{112}{15} = -299,88$

b) För att bilda ett 95 %-igt prediktionsintervall måste det gälla att observationerna på y är normalfördelade och stokastiskt oberoende och att de har samma varians. Residualerna beräknas genom $e_i = y_i - (-299,88 + 54,93 x_i)$ och residualkvadratsumman blir $SS_E = \sum e_i^2 = 1464,051$. Detta ger oss $s_e^2 = SS_E / (n-2) = 1464,051 / (15-2) = 112,619$. En uppskattning av μ_x då $x = 7$ blir $\widehat{\mu}_7 = -299.88 + 54.93 \cdot 7 = 84.63$. Variansen för en prediktion av μ_7 är $s_0^2 = s_e^2 \left(1 + \frac{1}{n} + \frac{(7-\overline{x})^2}{\sum (x_i - \overline{x})^2}\right) = 112.619 \left(1 + \frac{1}{15} + \frac{(7 - 112/15)^2}{846.26 - 112^2/15}\right) = 122,58$. Ett 95 %-igt prediktionsintervall då $x = 7$ ges nu av $\widehat{\mu}_7 \pm t_{0,975}(n-2)\, s_0$. Eftersom $t_{0,975}(15-2) = 2,16$ får vi prediktionsintervallet $84,63 \pm 2.16\sqrt{122.58}$, dvs $84,63 \pm 23,915$.

28.13

a) Låt x vara ålder och y vara tallplantans höjd. Beteckna väntevärdet av y med μ_x. Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



Från observationerna beräknar vi först $\sum x_i = 28$, $\sum x_i^2 = 140$, $\sum y_i = 253$, $\sum y_i^2 = 11599$, $\sum x_i y_i = 1271$ och $n = 7$. Detta ger uppskattningarna

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{1271 - 28 \cdot 253/7}{140 - 28^2/7} = 9,25$

$a = \overline{y} - b\overline{x} = \frac{253}{7} - 9.25 \cdot \frac{28}{7} = -0,85714$

b) Vi ska nu pröva om hypotesen att lutningsparametern är 0. För detta måste det gälla att observationerna på y är normalfördelade med samma varians och att de är stokastiskt oberoende. Vi beräknar först residualerna $e_i = y_i - (-0,85714 - 9,25x_i)$ och residualkvadratsumman $SS_E = \sum e_i^2 = 59,107$ och residualvariansen är $s_e^2 = SS_E / (n-2) = 59,107 / (7-2) = 11,821$. Variansen av uppskattningen av β är därför $s_b^2 = s_e^2 / \sum (x_i - \overline{x})^2 = 11.821 / (140 - 28^2/7) = 0,42218$. Observerat värde på testvariabeln blir därför $t_{obs} = b/s_b = 9.25/\sqrt{0.42218} = 14,236$. Hypotesen om ingen lutning förkastas på 5 %-nivån om $|t_{obs}| > t_{0,975}(n-2) = 2,571$, dvs. hypotesen om ingen lutning frkastas på 5%-nivån.

Alternativt kan hypotesen testas genom att beräkna $SS_T = \sum (y_i - \overline{y})^2 = 11599 - 253^2/7 = 2454,9$ och testvariabeln $F_{obs} = \frac{SS_T - SS_e}{SS_E/(n-2)} = \frac{2454.9 - 59.107}{59.107/(7-2)} = 202,67$. Hypotesen förkastas på 5%-nivån om $F_{obs} > F_{0,95}(1, n-2) = 6,61$, dvs. vi kan förkasta hypotesen om ingen lutning.

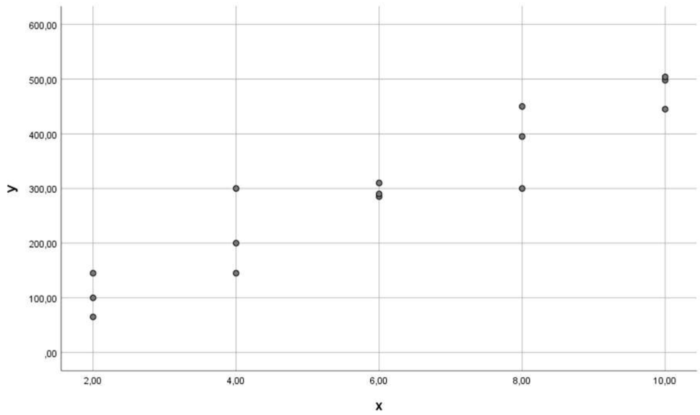
c) En uppskattning av μ_x då $x = 4$ är $\widehat{\mu}_4 = -0.85714 + 9.25 \cdot 4 = 36.143$.

Variansen för en prediktion av μ_4 är $s_0^2 = s_e^2 \left(1 + \frac{1}{n} + \frac{(4-\overline{x})^2}{\sum(x_i - \overline{x})^2}\right)$

$= 11.821 \left(1 + \frac{1}{7} + \frac{(4-28/7)^2}{140-28^2/7}\right) = 13,510$. Ett 95 %-igt prediktionsintervall då $x = 4$ ges därför av $\widehat{\mu}_4 \pm t_{0,975}(n-2)\, s_0$. Eftersom $t_{0,975}(5-2) = 2,571$ får vi prediktionsintervallet $36,143 \pm 2.571\sqrt{13.510}$, dvs $36,143 \pm 9,45$.

28.14

a) Låt x vara exponeringsutrymme och y vara försäljning i kr. Beteckna väntevärdet av y med μ_x. Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



Från observationerna beräknar vi först $\sum x_i = 90$, $\sum x_i^2 = 660$, $\sum y_i = 4432$, $\sum y_i^2 = 1596279$, $\sum x_i y_i = 32140$ och $n = 15$. Vi får då uppskattningarna

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{32140 - 90 \cdot 4432/15}{660 - 90^2/15} = 46,233$

$a = \overline{y} - b\overline{x} = \frac{4432}{15} - 46.233 \cdot \frac{90}{15} = 18,069$

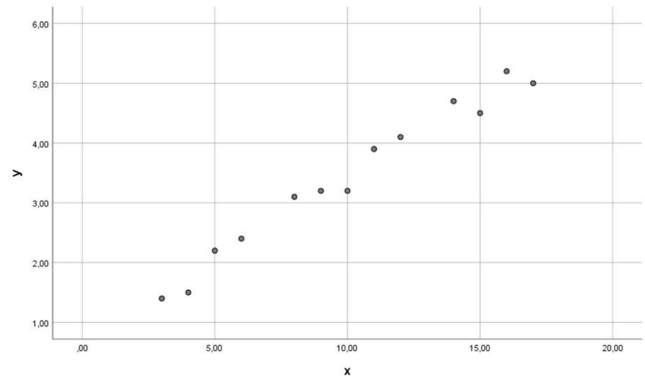
b) För att beräkna förklaringsgraden beräknar vi först residualerna $e_i = y_i - (18,069 - 46,233 x_i)$ och residualkvadratsumman $SS_E = \sum e_i = 30259,2$. Totala kvadratsumman är $SS_T = \sum (y_i - \overline{y})^2 = 1596279 - 4432^2/15 = 286761,733$. Förklaringsgraden beräknas då till $R^2 = \frac{SS_T - SS_E}{SS_T} = \frac{286761.733 - 30259.2}{286761.733} = 0,894$

c) En uppskattning av variansen för uppskattningen av β ges av $s_b^2 = s_e^2 / \sum (x_i - \overline{x})^2 = (30259.2/(15-2))/(660-90^2/15) = 19,397$. Ett konfudensintervall för β är $b \pm t_{0,975}(n-2)\, s_b$,dvs. $46,233 \pm 2.16 \cdot \sqrt{19.397}$ som uträknat blir $46,233 \pm 9,5131$. Eftersom ett 95%-igt konfidensintervall för β inte täcker 0, kan vi förkasta nollhypotesen $H_0 : \beta = 0$ med signifikansnivån 5%.

d) En uppskattning av μ_x då $x = 8$ är $\widehat{\mu}_8 = 18.069 + 46.233 \cdot 8 = 387,93$. Vidare får vi att en uppskattning av variansen för $\widehat{\mu}_8$ är $s_0^2 = s_e^2 \left(1 + \frac{1}{n} + \frac{(8-\overline{x})^2}{\sum (x_i - \overline{x})^2}\right) = \frac{30259.2}{15-2}\left(1 + \frac{1}{15} + \frac{(8-90/15)^2}{660-90^2/15}\right) = 2560,4$. Ett 95%-igt prediktionsintervall då $x = 8$ är $\widehat{\mu}_8 \pm t_{0,975}(n-2)\, s_0$ vilket ger $387,93 \pm 2.16 \cdot \sqrt{2560.4}$, dvs. $387,93 \pm 109,30$.

28.15

a) Låt x vara antal dagar efter kläckning och y vara vinglängden. Beteckna väntevärdet av y med μ_x. Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



Från observationerna beräknar vi först $\sum x_i = 130$, $\sum x_i^2 = 1562$, $\sum y_i = 44,40$, $\sum y_i^2 = 171,30$, $\sum x_i y_i = 514,80$ och $n = 13$. Detta ger uppskattningarna

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{514.80 - 130 \cdot 44.40/13}{1562 - 130^2/13} = 0,27023$

$a = \overline{y} - b\overline{x} = \frac{44.40}{13} - 0.27023 \cdot \frac{130}{13} = 0,71308$

b) För att beräkna observationernas varians runt linjen beräknar vi först residualerna $e_i = y_i - (0,71308 + 0,27023 \cdot x_i)$ och sedan $SS_E = \sum e_i^2 = 0,525$. En uppskattning av observationernas varians runt linjen blir därmed $s_e^2 = SS_E / (n-2) = 0,525 / (13-2) = 0,048$. Vidare får vi $SS_T = \sum (y_i - \overline{y})^2 = 171.30 - 44.4^2/13 = 19,657$ så att förklaringsgraden är $R^2 = \frac{SS_T - SS_E}{SS_T} = \frac{19.657 - 0.525}{19.657} = 0,973$. Eftersom det är endast en förklaringsvariabel i regressionen blir korrelationen mellan x och y lika med $R = \sqrt{0.973} = 0,986$. Korrelationen är positiv eftersom lutningskoefficienten b är positiv.

c) En uppskattning av variansen för b är $s_b^2 = s_e^2 / \sum (x_i - \overline{x})^2 = 0.048 / (1562 - 130^2/13)$ $0,00018321$. Ett 95%-igt konfidensintervall för β blir därmed $b \pm t_{0,975}(n-1)\, s_b$, dvs. $0,2702 \pm 2.201 \cdot \sqrt{0.00018321}$ eller $0,2702 \pm 0,0298$. Eftersom det 95%-iga konfidensintervallet inte täcker 0 förkastas hypotesen att lutningsparametern är 0 på signifikansnivån 5 %

d) En uppskattning av μ_x då $x = 10$ är $\widehat{\mu}_{10} = 0.71308 + 0.27023 \cdot 10 = 3,4154$. Vidare får vi att en uppskattning av variansen för $\widehat{\mu}_{10}$ är $s_0^2 = s_e^2 \left(1 + \frac{1}{n} + \frac{(10-\overline{x})^2}{\sum (x_i - \overline{x})^2}\right) = 0.048 \left(1 + \frac{1}{13} + \frac{(10-130/13)^2}{1562 - 130^2/13}\right) = 0,052$. Ett 95%-igt prediktionsintervall då $x = 10$ är $\widehat{\mu}_{10} \pm t_{0,975}(n-2)\, s_0$ vilket ger $3,4154 \pm 2.201 \cdot \sqrt{0.052}$, dvs. $3,4154 \pm 0,5019$.

28.16

Vi vet att $n = 7$, $SS_R = 15,36$ och $F = 16$. Eftersom $SS_R = SS_T - SS_E$ och $F = \frac{SS_T - SS_E}{SS_E/(n-2)}$ betyder det att $16 = \frac{15,36}{SS_E/(7-2)}$, dvs. $SS_E = 5\frac{15,36}{16} = 4,8$. Observationernas varians runt regressionslinjen uppskattas med $s^2 = \frac{SS_E}{n-2} = \frac{4,8}{7-2} = 0,96$ Vidare får vi att $SS_T = SS_R + SS_E = 15,36 + 4,8 = 20,16$ så att förklaringsgraden är $R^2 = \frac{SS_R}{SS_T} = \frac{15,36}{20,16} = 0,76$. För att pröva $H_0 : \beta = 0$ använder vi testvariabeln F och förkastar H_0 på signifikansnivån 5 procent om $F > F_{0,95}(1, n-2) = 6,61$. Vi ser att det observerade värdet på F överstuger $6,61$ så att H_0 förkastas på 5 procents nivån.

28.17

a) Låt x vara kostnad för marknadsföring och y vara försäljning. Beteckna väntevärdet av y med μ_x. Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Från uppgiften vet vi att $\sum x_i = 33$, $\sum x_i^2 = 132$, $\sum y_i = 377$, $\sum y_i^2 = 14722$, $\sum x_i y_i = 1314$ och $n = 10$. Detta ger uppskattningarna

$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{1314 - 33 \cdot 377/10}{132 - 33^2/10} = 3,0260$

$a = \overline{y} - b\overline{x} = \frac{377}{10} - 3.0260 \cdot \frac{33}{10} = 27,714$

b) Vi noterar först att $SS_E = \sum \left(y_i - (a + bx_i)\right)^2 = \sum \left(y_i - (\overline{y} - b\overline{x} + bx_i)\right)^2 = \sum \left((y_i - \overline{y}) + b\left(\overline{x} - x_i\right)\right)^2$

$= \sum \left((y_i - \overline{y})^2 - 2b\left(x_x - \overline{x}\right)\left(y_i - \overline{y}\right) + b^2\left(x_i - \overline{x}\right)^2\right)$

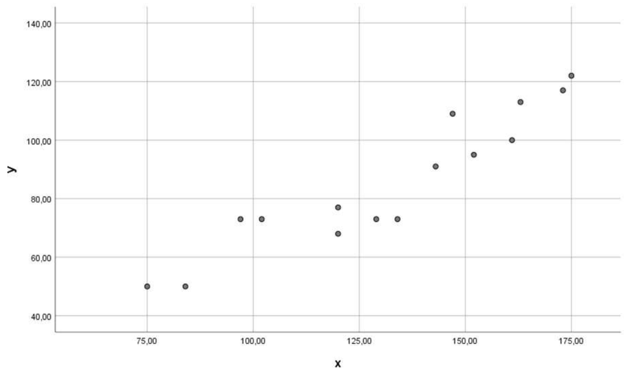
$= \sum \left(y_i - \overline{y}\right)^2 - 2\frac{\sum (x_x - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} \sum \left(x_x - \overline{x}\right)\left(y_i - \overline{y}\right) + \left(\frac{\sum (x_x - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}\right)^2 \sum \left(x_i - \overline{x}\right)^2 = \sum \left(y_i - \overline{y}\right)^2 - \frac{\sum (x_x - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} \sum \left(x_x - \overline{x}\right)\left(y_i - \overline{y}\right)$

$= SS_T - b \sum \left(x_x - \overline{x}\right)\left(y_i - \overline{y}\right)$.

I vårt fall får vi att $SS_T = 14722 - 377^2/10 = 509,1$ och $SS_E = 509.1 - 3.0260 \cdot (1314 - 33 \cdot 377/10) = 297,58$. För att pröva hypotesen H_0-$\beta = 0$ använder vi testvariabeln $F = \left(SS_T - SS_E\right) / \left(SS_E / \left(n - 2\right)\right)$ och förkastar H_0 på signifikansnivån 5% om $F_{obs} > F_{0,95}\left(1, n - 2\right)$. I vårt fall får gäller det att $F_{0,95}\left(1, n - 2\right) = 5.32$ och $F_{obs} = \left(509 - 297.58\right) / \left(297.58 / \left(10 - 2\right)\right) = 5,6837$, dvs. H_0 förkastas.

28.18

Låt x vara vikten hos ett äpple och y vara mängden juice. Beteckna väntevärdet av y med μ_x. Ett linjärt samband mellan x och y är $\mu_x = \alpha + \beta x$. Ett punktdiagram över observationerna är



Från observationerna beräknar vi först $\sum x_i = 1975$, $\sum x_i^2 = 274097$, $\sum y_i = 1284$, $\sum y_i^2 = 117398$, $\sum x_i y_i = 178662$ och $n = 15$. Detta ger uppskattningarna

$$b = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \left(\sum x_i\right)^2 / n} = \frac{178662 - 1975 \cdot 1284/15}{274097 - 1975^2/15} = 0,68316$$

$$a = \overline{y} - b\overline{x} = \frac{1284}{15} - 0.68316 \cdot \frac{1975}{15} = -4,3494$$

För att beräkna modellens förklaringsgrad beräknar vi först residualerna $e_i = y_i - (-4,3494 + 0,68316 \cdot x_i)$ och residualkvadratsumman $SS_E = \sum e_i^2 = 927,926$. Vidare får vi att $SS_T = 117398 - 1284^2/15 = 7487,6$ så att $R^2 = (7487.6 - 927.926)/7487.6 = 0,876$. Vi får även att $s_e^2 = SS_E/(n-2) = 927.926/(15-2) = 71.379$. Uppskattningar av varianserna för a respektive b är $s_a^2 = s_e^2\left(\frac{1}{n} + \frac{\overline{x}^2}{(n-1)s_x^2}\right) = 71.379\left(\frac{1}{15} + \frac{(1975/15)^2}{274097-1975^2/15}\right) = 92,799$ och $s_b^2 = s_e^2/\sum(x_i - \overline{x})^2 = 71.397/(274097 - 1975^2/15) = 0.0050797$. 95%-iga konfidensintervall för parametrarna ges därför av $a \pm t_{0,975}(n-1)\, s_a$ respektive $b \pm t_{0,975}(n-1)\, s_b$. Vi får för α $4,3494 \pm 2.145\sqrt{92.799}$, dvs. $4,3494 \pm 20,663$ och för β $0,68316 \pm 2.145\sqrt{0.0050797}$, dvs. $0,6832 \pm 0,1529$.

29. LINJÄR REGRESSION MED FLERA FÖRKLARINGSVARIABLER

29.1

Låt x_1 vara gödselmängd, x_2 regnmängd och y skörd och beteckna förväntat värde av y med $\mu_{\mathbf{X}}$. En linjär regressionsmodell är nu $\mu_{\mathbf{X}} = \alpha + \beta_1 x_1 + \beta_2 x_2$. Om data analyseras med ett datorprogram för linjär regression får vi minstakvadratskattningarna $a = 3057,783$, $b_1 = 3,627$ och $b_2 = -14,359$. Skattningarna kan givetvis även fås genom att lösa ekvationssystemet på sidan 545. t-värden för parameterskattningarna är för a $t_{obs} = 13,821$, dvs. klart signifikant. Däremot är tolkningen av α inte helt meningsfull (skörd utan gödsel och inget regn). För β_1 och β_2 är motsvarande t-värden $26,273$ (tydligt signifikant) respektive $-5,723$ (också tydligt signifikant). Kvadratsummorma sammanfattas i ev ANOVA tabell:

Variation	SS	fg	MS
Regression	4213339,092	2	2106669,546
Residual	20310,908	4	5077,727
Total	4233650,000	6	

En uppskattning av observationernas varians är $SS_E/(n-3) = 20310,908/(7-3) = 5077727$. Förklaringsgraden är $R^2 = SS_R/SS_T = 4213339,092/4233650 = 0,995$.

29.2

Låt x_1 vara aktiv lungkapacitet, x_2 total lungkapacitet och y utandningskapacitet. Beteckna förväntat värde av y med $\mu_{\mathbf{X}}$. En linjär regressionsmodell är $\mu_{\mathbf{X}} = \alpha + \beta_1 x_1 + \beta_2 x_2$. Om data analyseras med ett datorprogram för linjär regression får vi minstakvadratskattningarna $a = -0,159$, $b_1 = 0,747$ och $b_2 = 0,037$. Skattningarna kan givetvis även fås genom att lösa ekvationssystemet på sidan 545. t-värden för parameterskattningarna för lutningsparametrarna är, för β_1 lika med $10,391$, dvs. signifikant på 5 %-nivån, och för β_2 lika med $0,491$, dvs. inte signifikant på 5 %-nivån. Förklaringsgraden är $R^2 = 0,939$. En uppskattning av förväntad lungkapacitet med aktiv lungkapacitet $x_1 = 2$ och total lungkapacitet $x_2 = 4$ är $\widehat{\mu}_{2,4} = 0.159 + 0.747 \cdot 2 + 0.037 \cdot 4 = 1,801$.

29.3

Vi vet att $n = 20$ och att $SS_T = 116300$. Om både x_1 och x_2 är med i modellen får vi residualkvadratsumman $SS_E = 34250$. Vi vill nu pröva om var och en av förklaringsvariablerna påverkar väntevärdet för y, dvs. vi vill pröva om motsvarande lutningsparameter är noll. Testet blir beroende av om den andta förklaringsvariabeln antas vara med i modellen eller inte. Vi beskriver här fallet då den andra förklaringsvariabeln inte är med i modellen. För att pröva hypotesen $H_0 : \beta_1 = 0$ noterar vi att residualkvadratsumman om endast x_1 är förklaringsvariabel är $SS_E = 108100$, dvs. vi får teststatistikan $F_{obs} = (116300 - 108100) / (116300 / (20 - 2)) = 1,2691$. H_0 förkastas på 5%-nivån om $F_{obs} > F_{0,95}(1, 20 - 2) = 4,41$, dvs. H_0 kan inte förkastas. För att pröva hypotesen $H_0 : \beta_2 = 0$ noterar vi att residualkvadratsumman om endast x_2 är förklaringsvariabel är $SS_E = 35760$, dvs. vi får teststatistikan $F_{obs} = (116300 - 35760) / (116300 / (20 - 2)) = 12,465$, dvs. vi har empiriskt evidens för att x_2 påverkar väntevärdet för y.

Förklaringsgraden för modellen men bara x_1 som förklaringsvariabel är $R^2 =$

$(116300 - 108100) / 116300 = 0,07$, om bara x_2 används blir det $R^2 =$

$(116300 - 35760) / 116300 = 0,69$ och om båda förklaringsvariablerna används i modellen blir förklaringsgraden $R^2 = (116300 - 34250) / 116300 = 0,705$

29.4

Vi inför beteckningarna $x_1 =$ BMI, $x_2 =$ ålder, $x_3 = 1$ om personen är rökare och 0 annars och $y =$ systoliskt blodtryck. Beteckna förväntat värde av y med $\mu_{\mathbf{X}}$. En linjär regressionsmodell är $\mu_{\mathbf{X}} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Om data analyseras med ett datorprogram för linjär regression får vi minstakvadratskattningarna $a = 32,671$, $b_1 = 0,629$, $b_2 = 1,365$ och $b_3 = 6,874$. Skattningarna kan givetvis även fås genom att lösa ekvationssystemet på sidan 545. t-värdet för parameterskattningarna för lutningsparametrarna är, för β_1 lika med $3,643$, dvs. β_1 är signifikant skild från 0 på 5 %-nivån, för β_2 är t-värdet lika med $8,512$, vilket innebär att också β_2 är signifikant skild från 0 på 5 %-nivån. t-värdet för β_3 är 3,695 vilket innebär att β_3 är signifikant skild från moll. Vi sammanfattar kvadratsummorna i följande tabell

Variation	SS	fg	MS
Regression	2395,496	3	798,499
Residual	257,704	16	16,106
Total	2653,200	19	

Observationernas varians kring regressionen är $s^2 = MS_E = SS_E / (n-4) = 257,704 / (20-4) = 16,106$. Förklaringsgraden är $R^2 = SS_R/SS_T = 2395,496/2653,200 = 0,903$.

29.5

Vi prövar först hypotesen att det inte finns någon skillnad i blodtryck mellen män och kvinnor. Vi använder ett t-test för två orelsterade urval. I urvalet med män får vi urvalsmedelvärdet $\overline{y}_M = 156,9$ och urvalsvariansen $s_M^2 = 262,31$. I urvalet med kvinnor får vi urvalsmedelvärdet $\overline{y}_K = 144,85$ och urvalsvariansen $s_K^2 = 642,34$. Den poolade variansen blir då $s_p^2 = \frac{(20-1)\cdot 262.32 + (20-1)\cdot 642.34}{20+20-2} = 452,33$. Det observerade t-värdet kan då beräknas till $t_{obs} = (\overline{y}_M - \overline{y}_K) / \left(s_P \sqrt{\frac{1}{n_M} + \frac{1}{n_K}} \right) = (156.9 - 144.85) / \left(\sqrt{452.31} \sqrt{\left(\frac{1}{20} + \frac{1}{20} \right)} \right) = 1,792$. Vi ska jämföra det observerade t-värdet med percentiler i en t-fördelning med $20+20-2 = 38$ frihetsgrader. Eftersom antalet frihetsgrader är stort kan vi approximera fördelningen med en standard normalfördelning. Vi förkastar därför hypotesen om lika blodtryck för män och kvinnor om $|t_{obs}| > 1,96$. Tydligen ger inte våra data evidens för att det är någon skillnad i blodtryck mellan män och kvinnor.

Vi prövar nu hypotesen att det inte finns någon skillnad i blodtryck mellan män och kvinnor då vi tar hänsyn åldersfördelningen i de båda urvalen. Vi gör det genom att analysera en regressionsmodell med blodtryck som beroende variabel och ålder och kön som förklaringsvariabler. Här är kön en dummyvariabel och att pröva hypotesen om lika blodtryck bland män och kvinnor är ekvivalent med att pröva om parametern för kön är lika med noll. En regressionsmodell med endast älder som förklaringsvariabel ger residualkvadratsumman $SS_0 = 9658,284$. Om även kön är med i modellen som en förklaringsvariabel får vi residualkvadratsumman $SS_E = 8824,573$. Det observerade F-värdet blir då $(9658.284 - 8824.573) / (8824.573/ (40 - 3)) = 3,5$. Hypotesen om lika blodtryck för män och kvinnor förkastas på 5%-nivån om $F_{obs} > F_{0,95}(1,37) = 4,11$, dvs. vi kan inte förkasta hypotesen. Vi ser här att ålder verkar inte vara en konfounder som döljer en eventuell skillnad i blodtryck mellan män och kvinnor.

29.6

Vi beräknar först summor och kvadratsummor för de tre variablerna. För Inkomst får vi $\sum x_{I,i} = 225$ och $\sum x_{I,i}^2 = 9375$, för antal barn får vi $\sum x_{B,i} = 12$ och $\sum x_{B,i}^2 = 48$ och för konsumtion får vi $\sum x_{K,i} = 27,6$ och $\sum x_{K,i}^2 = 135,52$. Varianserna för de tre variablerna blir då $s_I^2 = (9375 - 225^2/6) / (6-1) = 187,5$, $s_B^2 = (48 - 12^2/6) / (6-1) = 4,8$ och $s_K^2 = (135.52 - 27.6^2/6) / (6-1) = 1,712$. För att kunna beräkna korrelationer behöver vi först beräkna kovarianserna. Produktsummorna är $\sum x_{I,i}x_{B,i} = 400$, $\sum x_{I,i}x_{K,i} = 1020$ och $\sum x_{B,i}x_{K,i} = 68,8$. Kovarianserna blir därmed $s_{I,B} = (400 - 225 \cdot 12/6) / (6-1) = -10,0$, $s_{I,K} = (1020 - 225 \cdot 27.6/6) / (6-1) = -3,0$ och $s_{B,K} =$

$(68.8 - 12 \cdot 27.6/6) / (6-1) = 2,72$.

Korrelationen mellan konsumtion och inkomst blir därmed $r_{I,K} = s_{I,K}/ (s_I s_K) = -3/\sqrt{187.5 \cdot 1.712} = -0,16744$, ett något förvånande resultat eftersom det innebär att familjer med högre inkomst konsumerar mindre.

Vi beräknar nu den partielle korrelationen mellan konsumtion och inkomst givet antal barn. För det behöver vi först beräkna korrelationerna mellan inkomst och barn, $r_{I,B} = s_{I,B}/ (s_I s_B) = -10/\sqrt{187.5 \cdot 4.8} = -0,33333$ och $r_{B,K} = s_{B,K}/ (s_B s_K) = 2.72/\sqrt{4.8 \cdot 1.712} = 0,94885$. Den partiella korrelationen mellan konsumtion och inkomst när vi tar hänsyn till antal barn blir därmed $r_{I,K \cdot B} = (r_{I,K} - r_{I,B} r_{K,B}) / \left(\sqrt{1 - r_{I,B}^2}\sqrt{1 - r_{K,B}^2}\right) = ((-0.16744) - (-0.33333) \cdot 0.94885) / \left(\sqrt{1 - 0.33333^2}\sqrt{1 - 0.94885^2}\right) = 0,5$, vilket är ett mer förväntat resultat. Vi ser att antal barn är en konfounder som påverkar sambandet mellan inkomst och konsumtion.

29.7

Låt x vara poäng på förtestet, y poäng på testet efter försöket och beteckna väntevärdet av y med μ_x. En regressionsmodell är då $\mu_x = \alpha + \beta x$. Parametrarna α och β kan uppskattas med ett datorprogram för regression. Detta ger skattningarna $a = 3,029$ och $b = 1,037$. Residualkvadratsumman är $SS_0 = 40,011$. Vi vill nu pröva om sambandet är lika i de båda grupperna och introducerar därför en dummyvariabel d som har värdet 0 för metod A och värdet 1 för metod B. Vi bildar också en ny variabel som probdukten av variablerna x och d, $xd = x \cdot d$. En regressionsmodell med dessa variabler är $\mu_x = \alpha + \beta_1 x + \beta_2 d + \beta_3 xd$. Om sambandet är lika i de båda grupperna är $\beta_2 = \beta_3 = 0$. För att pröva hypotesen att dessa parametrar är 0 skattar vi modellen med alla tre förklaringsvariablerna. Den ger residualkvadratsumman $SS_E = 28,741$. Det observerade F-värdet blir därför $F_{obs} = ((40.011 - 28.741)/2)/(28.741/(10-4)) = 1,1764$. Hypotesen om lika samband förkastas på 5%-nivån om $F_{obs} > F_{0,95}(2,6) = 5,14$, dvs. det finns inget evidens för att sambandet är olika i de två grupperna.

29.8

Vi ska uppskatta parametrar i en llinjär regressionsmodell med givna summor av observationerna. Det ekvationssystem som definierar parameterskattningarna är

$$\begin{cases} 20a & +38b_1 & +34b_2 & = & 189 \\ 38a & +154b_1 & +91b_2 & = & 673 \\ 34a & +91b_1 & +150b_2 & = & 579 \end{cases}$$

Ekvationssystemet kan t.ex.lösas med successiv eliminering och ger lösningen

$$a = 0,128$$
$$b_1 = 3,234$$
$$b_2 = 1,869$$

29.9

Den partiella korrelationen mellan vikten av ett fullvuxet träd, X, och skottens tillväxt under de första fyra åren, Y, då inverkan från stammens omfång vid marken hos ett fullvuxet träd, Z, är

$$r_{X,Y\cdot Z} = \frac{r_{X,Y} - r_{X,Z}\cdot r_{Y,Z}}{\sqrt{1-r_{X,Z}^2}\sqrt{1-r_{Y,Z}^2}} = \frac{0.2-0.9\cdot 0.4}{\sqrt{1-0.9^2}\sqrt{1-0.4^4}} = -0,3719$$

29.10

Låt P vara priset per kg på kalkon, T utbudet av kalkon per person och C utbudet av kyckling per person med observationer under sex år. Låt μ_P vara förväntat värde av P. En linjär regressionsmodell är då $\mu_P = \alpha + \beta_1 T + \beta_2 C$. För att uppskatta parametrarna sätter vi upp ekvationssystemet

$$\begin{cases} 6a & +14,9b_1 & +105b_2 & = & 554 \\ 14,9a & +37,89b_1 & +264,8b_2 & = & 1345,6 \\ 105a & +264,8b_1 & +1863b_2 & = & 9548 \end{cases}$$

med lösningen

$$a = 184,978$$
$$b_1 = -27,824$$
$$b_2 = -1,346$$

Om utbudet av kyckling (C) ökar med två enheter samtidigt som utbudet av kalkon (T) minskar med en halv enhet förväntar vi oss att priset på kalkon (P) förändras med $(-27.824)\cdot(-0.5)+(-1.345)\cdot 0.5 = 13,24$, dvs vi förväntar oss att priset ökar med $13,24$ kr per kg.

29.11

Låt Y vara draghållfastheten och X vara tjockleken på stärkelsefilm. Definiera en dummyvariabek d genom att den får värdet 0 om observationen är från Dasheen och 1 om den är från Sweet potato. En linjär regressionsmodell är $E(Y) = \alpha + \beta_1 X + \beta_2 d$. I denna modell antar vi att det är samma lutning för de två stärkelsefilmerna. Vi skattar modellen med hjälp av ett datorprogram för linjär regression och får residualkvadratsumman $6293,498$. Bilda nu en ny variabel, $xd = X \cdot d$ och skatta modellen $E(Y) = \alpha + \beta_1 X + \beta_2 d + \beta_3 xd$. I denna modell får lutningarna vara olika för de båda stärkelsefilmerna och parametern β_3 anger skillnaden i lutning. När denna modell skattas får vi residualkvadratsumman $SS_E = 655,333$. Genom att tillåta olika lutningar förbättras alltså modellanpasningan med $SS_0 = 6293.498 - 655.333 = 5638,2$. Vi kan nu testa om förändringen är signifikant med ett F-test och använder då testvariabeln $F = SS_0/\left(SS_E/\left(n-p\right)\right)$ och förkastar hypotesen om lika lutning på 5%-nivån om $F_{obs} > F_{0,95}\left(1, n-p\right) = F_{0,95}\left(1, 9-4\right) = 6,61$. Vi får att $F_{obs} = 5638.2/\left(655.333/\left(9-4\right)\right) = 43,018$, dvs. det finns empiriskt evidens för att lutningen är olika för de två stärkelsesorterna.

29.12

a) Vi vet att *df* för total är $100 - 1 = 99$ och för regression 5. Vi kan då komplettera tabellen enligt följande:

	SS	*df*	MS	F
Regression	822	5	164,4	82,2
Residual	188	94	2,0	
Total	1000	99		

b) $R^2 = SS_R/SS_T = 822/1000 = 0,822$

c) För att pröva hypotesen $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ mot att minst en av parametrarna är skild från 0, måste det gälla att observationerna på beroendevariabeln är normalfördelade med konstant varians, att alla observationer är stokastiskt oberoende av varandra och att förklaringsvariablerna är måtta utan mätfel. Testvariabel för att pröva H_0 är F och H_0 förkastas på 5%-nivån om $F_{obs} > F_{0,95}\left(5, 94\right) = 4,40$. Vi ser att $F_{obs} = 82,2$ så att H_0 förkastas på 5%-nivån.

29.13

a) Låt x_1 reklamkonstander i tv, x_2 vara reklamkonstnader i bio, x_3 reklamkostnader i tidningar och Y försäljning av marksten. Beteckna väntevärdet av Y med μ_x. En regressionsmodell är då $\mu_{\mathbf{x}} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Parametrarna α, β_1,β_2 och β_3 kan uppskattas med ett datorprogram för regression. Detta ger skattningarna $a = -8,473$, $b_1 = 5,158$, $b_2 = -0,797$ och $b_3 = 1,004$.

b) Förklaringsgraden är $R^2 = 0,964$.

c) För att pröva hypotesen $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ mot att någon β_j är skild från 0, måste det gälla att Y är normalfördelad med samma varians för alla observationer, att observationerna är stokastiskt oberoende och att det inte är några mätfel i förklaringsvariablerna. Vi använder F-testet och förkastar H_0 på 5%-nivån om $F_{obs} > F_{0{,}95}(3,8) = 4{,}07$. Vi får att $F_{obs} = ((SS_T - SS_E)/3)/(SS_E/(n-p)) = ((5174.917 - 187.136)/3)/(187.136/(12-4)) = 71{,}075$, dvs. H_0 förkastas på 5%-nivån.

29.14

a) För att pröva hypotesen $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ mot att någon β_j är skild från 0, måste det gälla att Y är normalfördelad med samma varians för alla observationer, att observationerna är stokastiskt oberoende och att det inte är några mätfel i förklaringsvariablerna. Vi använder F-testet och förkastar H_0 på 5%-nivån om $F_{obs} > F_{0{,}95}(3,12) = 3{,}49$. Vi får att $F_{obs} = ((SS_T - SS_E)/3)/(SS_E/(n-p)) = ((1280000 - 84832)/3)/(84832/(16-4)) = 56{,}355$, dvs. H_0 förkastas på 5%-nivån.

b) Vi ska nu pröva om var och en av parametrarna är signifikant skilld från 0. Vi använder t-testet och förkastar $H_0 : \beta_j = 0$ mot $H_A : \beta_j \neq 0$ på 5%-nivån om $|t_{obs}| > t_{0{,}975}(n-p) = 2{,}179$ för $j = 1,2,3$. Observerade t-värden fås genom $t_{obs} = b_j/s_{b_j}$. För inkomst får vi $t_{obs} = 0.12/0.0165 = 7{,}2727$, för banktillgångar får vi $t_{obs} = -0.026/0.0060 = -4{,}3333$ och för antal barn får vi $t_{obs} = -22.76/23.49 = -0.96892$, dvs. inkomst och banktillgångar är signifikanta på 5%-nivån medan antal barn inte är det.

c) En uppskattning av variansen är $s^2 = SS_E/(n-p) = 84832/(16-4) = 7069{,}3$. Modellens förklaringsgrad är $R^2 = (SS_T - SS_E)/SS_E =$

$(1280000 - 84832)/1280000 = 0{,}93$.

30. ANALYS AV TIDSSERIER

30.1

Data och beräkningar visas i tabellerna nedan.

För att beräkna säsongkomponenterna bildar först fyratermers löpande medelvärden, dvs. för en tidpunkt mellan andra och tredje kvartalet 2010 får vi $\frac{89.9+102.4+98.1+109.6}{4} = 100,0$, osv.

Centrera de löpande medelvärdena så att för tredje kvartalet 2010 får det utjämnade värdet $\frac{100.0+102.775}{2} = 101,39$

Tidpunkt	Omsättning	4 term löp mv	Centrerat	Säsongrensad
2010 I	89,9			94,88
2010 II	102,4			100,01
2010 III	98,1	100,000	101,39	99,98
2010 IV	109,6	102,775	103,91	105,13
2011 I	101,0	105,050	106,26	105,98
2011 II	111,5	107,475	108,13	109,11
2011 III	107,8	108,775	109,61	109,68
2011 IV	114,8	110,450	109,89	110,33
2012 I	107,7	109,325	108,68	112,68
2012 II	107,0	108,025	107,14	104,61
2012 III	102,6	106,250	105,05	104,48
2012 IV	107,7	103,875	103,78	103,23
2013,I	98,2	103,675	103,61	103,18
2013,II	106,2	103,050	103,71	103,81
2013,III	102,1	103,875	104,16	103,98
2013,IV	109,0	104,450	105,26	104,53
2014,I	100,5	106,075	107,1	105,48
2014,II	112,7	108,125	108,85	110,31
2014,III	110,3	109,575	110,08	112,18
2014, IV	114,8	110,575	110,71	110,33
2015,I	104,5	110,850	111,13	109,48
2015,II	113,8	111,400		111,41
2015,III	112,5			114,38

Uppskattningar av säsongkomponenten fås nu genom att bilda differenser för observerade värden och utjämnade värden. För det tredje kvartalet 2010 får vi $98.1 - 101.39 = -3,29$ osv. för de andra kvartalen. Uppskattningarna förs in i tabellen nedan.

År	Kvartal			
	I	II	III	IV
2010			-3,29	5,69
2011	-5,26	3,37	-1,81	4,91
2012	-0,98	-0.14	-2,45	3,92
2013	-5,41	2,49	-2,06	3,74
2014	-6,60	3,85	0,22	4,09
2015	-6,63			
Medelvärde	-4,976	2,393	-1,878	4,470
Säsongkomponent	-4,978	2,391	-1,880	4,468

Bilda sedan medelvärden för varje kvartal. För kvartal I får vi $\frac{-5{,}26-0{,}98-5{,}41-6{,}60-6{,}63}{5} = -4,976$ och på samma sätt för de andra kvartalen.

För att uppskattningarna av säsongkomponenten ska summera till 0 subtraherar vi medelvärdet av dessa uppskattningar, dvs. vi subtraherar $\frac{-4{,}976+2{,}393-1{,}878+4{,}470}{4}$

$= 0,00225$. För första kvartalet blir därför uppskattningen av säsongkomponenten $-4,976 - 0,002 = -4,978$ och motsvarande för de andra kvartalen.

Säsongrensade värden fås nu genom att subtrahera säsongkomponenterna från de observerade värdena. För första kvartalet 2010 får vi $89.9-(-4.978) \approx 94,88$ osv. Resultaten är redovisade i kolumnen till höger i tabellen ovan. De säsongrensade värdena förmodas endast bero av trend och konjunktur.

30.2

a) Resultat från alla beräkningar visas i tabellerna nedan.

För att beräkna säsongkomponenterna bildar först fyratermers löpande medelvärden, dvs. för en tidpunkt mellan andra och tredje kvartalet 2017 får vi $\frac{99+63+73+167}{4} = 100,50$, osv.

Centrera de löpande medelvärdena så att för tredje kvartalet 2017 får det utjämnade värdet $\frac{100{,}50+102{,}75}{2} = 101,625$

Tidpunkt	Produktion	4 term löp mv	Centrerat	Säsongrensad	Exp. utj.
2017 I	99			100,44	99,00
2017 II	63			100,19	95,40
2017 III	73	100,50	101,625	92,06	93,16
2017 IV	167	102,75	104,375	109,31	100,54
2018 I	108	106,00	110,000	109,44	101,29
2018 II	76	114,00	114,875	113,19	98,76
2018 III	105	115,75	119,625	124,06	99,38
2018 IV	174	123,50	126,375	116,31	106,85
2019 I	139	129,25	133,000	140,44	110,06
2019 II	99	136,75	139,625	136,19	108,96
2019 III	135	142,50		154,06	111,56
2019 IV	197			139,31	120,10

Uppskattningar av säsongkomponenten fås nu genom att bilda differenser för observerade värden och utjämnade värden, Se tabellen nedan. För det tredje kvartalet 2017 får vi $73 - 101,625 = -28,625$ osv. för de andra kvartalen.

Bilda sedan medelvärden av differenserna för varje kvartal. För kvartal I får vi $\frac{-2,000-6,000}{2} = -4,000$ och på samma sätt för de andra kvartalen.

För att uppskattningarna av säsongkomponenten ska summera till 0 subtraherar vi medelvärdet av dessa uppskattningar, dvs. vi subtraherar $\frac{-4,000-39,750-21,625+55,12}{4}$

$= -2,5625$. För första kvartalet blir därför uppskattningen av säsongkomponenten $-4,000 - (-2,5625) = -1,4375$ och motsvarande för de andra kvartalen.

År	Kvartal			
	I	II	III	IV
2017			-28.625	62,625
2018	-2,000	-38,875	-14,625	47,625
2019	-6,000	-40,625		
Medelvärde	-4,000	-39,750	-21,625	55,125
Säsongkomponent	-1,438	-37,188	-19,063	57,688

b) Säsongkomponenterna subtraheras från de observerade värdena. För första kvartalet 2017 får vi $99 - (-1,438) \approx 100,44$ osv. Resultaten är redovisade i kolumnen till höger i tabellen ovan. De säsongrensade värdena förmodas endast bero av trend och konjunktur.

c) Med $\alpha = 0,1$ får vi den exponentiellt utjämnade serien som visas i den högra kolumnen i tabellen ovan. För t.ex. det andra kvartalet 2017 får vi $0,1 \cdot 63 + 0,9 \cdot 99 = 95,4$ osv. vilket ger $SS = 22686$.

30.3

a) För att beräkna säsongkomponenterna bildar först fyratermers löpande medelvärden, dvs. för en tidpunkt mellan andra och tredje kvartalet 2016 får vi $\frac{7.5+1.5+2.5+5.0}{4} = 4,125$, osv.

Centrera de löpande medelvärdena så att för tredje kvartalet 2016 får det utjämnade värdet $\frac{4,125+3,875}{2} = 4,0$

Tidpunkt	Elförbr.	4 term löp mv	Centrerat	Säsongrens	Exp.utj.	Holtz-Wint
2016 I	7,5			5,02		
2016 II	1,5			3,75	6,00	1,50
2016 III	2,5	4,125	4,0000	4,38	5,12	1,80
2016 IV	5,0	3,875	3,9375	3,73	5,09	4,33
2017 I	6,5	4,000	4,0625	4,02	5,44	6,18
2017 II	2,0	4,125	4,3125	4,25	4,58	2,43
2017 III	3,0	4,500	4,7500	4,88	4,19	2,80
2017 IV	6,5	5,000	5,2500	5,23	4,77	6,06
2018 I	8,5	5,500	5,6250	6,02	5,70	8,34
2018 II	4,0	5,750	5,9375	6,25	5,27	4,58
2018 III	4,0	6,125	6,1875	5,88	4,96	3,99
2018 IV	8,0	6,250	6,3125	6,73	5,72	7,54
2019 I	9,0	6,375	6,5625	6,52	6,54	8,96
2019 II	4,5	6,750	6,8125	6,75	6,03	5,07
2019 III	5,5	6,875		7.38	5,90	5,37
2019 IV	8,5			7,23	6,55	8,15

Uppskattningar av säsongkomponenten fås nu genom att bilda differenser för observerade värden och utjämnade värden, Se tabellen nedan. För det tredje kvartalet 2016 får vi $2,5 - 4,0 = -1,5$ osv. för de andra kvartalen.

Bilda sedan medelvärden av differenserna får varje kvartal. För kvartal I får vi $\frac{1,9375+2,8750+2,4375}{3} = 2,416\,7$ och på samma sätt för de andra kvartalen.

För att uppskattningarna av säsongkomponenten ska summera till 0 subtraherar vi medelvärdet av dessa uppskattningar, dvs. vi subtraherar $\frac{2,4167-2,1875-1,8125+1,333}{4}$

$= -0,0625$ För första kvartalet blir därför uppskattningen av säsongkomponenten $2,4167 - (-0,0625) = 2,4792$ och motsvarande för de andra kvartalen.

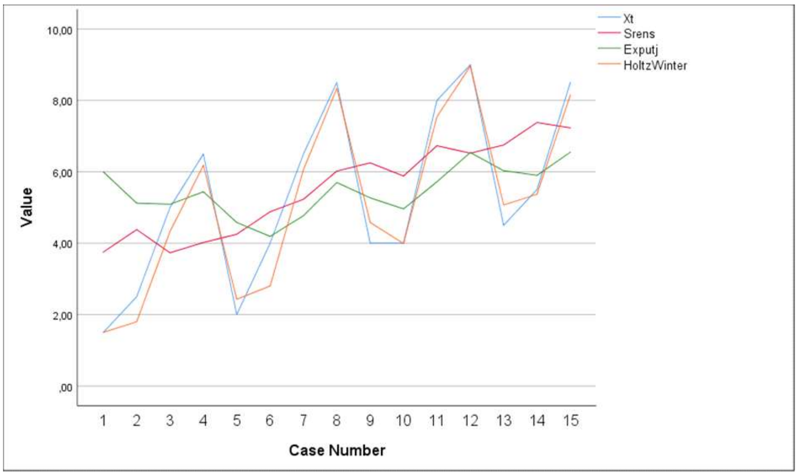
År	Kvartal			
	I	II	III	IV
2016			-1,5000	1,0625
2017	1,9375	-2,3125	-1,7500	1,2500
2018	2,8750	-1,9375	-2,1875	1,6875
2019	2,4375	-2,3125		
Medelvärde	2,4167	-2,1875	-1,8125	1,3333
Säsongkomponent	2,4792	-2,2500	-1,8750	1,2708

b) Säsongkomponenterna subtraheras från de observerade värdena. För första kvartalet 2016 får vi $7,5 - 2,4792 = 5,0208$ osv. Resultaten är redovisade i kolumnen till höger i tabellen ovan. De säsongrensade värdena förmodas endast bero av trend och konjunktur.

c) Med $\alpha = 0,25$ får vi den exponentiellt utjämnade serien som visas i den högra kolumnen i tabellen ovan. För t.ex. det andra kvartalet 2016 får vi $0,25 \cdot 1,5 + 0,75 \cdot 7,5 = 6,0$ osv. vilket ger $SS = 119,68$.

e) Prova med några olika värden på α och β, Värdena $\alpha = 0,9$ och $\beta = 0,4$ med initialvärdena $x^*_{16II} = x_{16II} = 1,5$ och $Tr^*_{16II} = x_{16II} - x_{16I} = 1,5 - 7,5 = -6$ ger $x^*_{16III} = 0,9 \cdot 2,5 + 0,1 \cdot (1,5 + (-6)) = 1,8$ och $Tr^*_{16III} = 0,4 \cdot (1,8 - 1,5) + 0,6 \cdot (-6) = -3,48$. Värdet på SS är $249,32$. Den utjämnade serien visas i tabellen ovan.

Seriens värden, den säsongrensade serien, den exponentiellt utjämnade serien och bearbetad med Holt-Winters metod visas i figuren nedan.



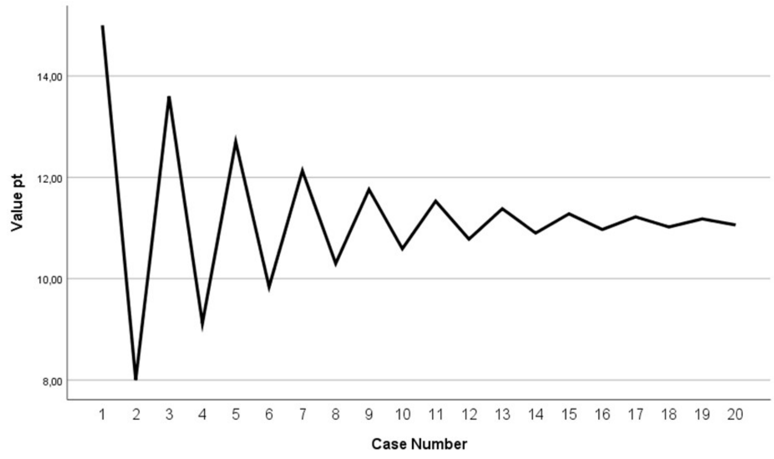
30.4

Sätt tillgång och efterfrågan lika med varandra: $d_0 - ap_t = s_0 + bp_{t-1}$ och lös ut priset vid tidpunkt t: $p_t = d_0 - s_0 - bp_{t-1} = \rho p_{t-1} + c$, där vi har gjort förenklingen $\rho = -b/a$ och $c = (d_0 - s_0)/a$. Notera att p_t är en första ordningens autoregressiv process.

Uttrycket för p_t ger oss nu att $p_{t.1} = \rho p_{t-2} + c$ Skriv in detta i uttrycket för p_t så att $p_t = \rho p_{t-1} + c = \rho(\rho p_{t-2} + c) + c = \rho^2 p_{t-2} + \rho c + c$. Fortsätt på samma sätt att skriva uttrcket för p_{t-2} och så vidare. Detta ger till slut uttrycket $p_t = \rho^t p_0 + \rho^{t-1} c + \rho^{t-2} c + \ldots + \rho^0 c = \rho^t p_0 + c \sum_{j=0}^{t-1} \rho^j$. Men detta är en geometrisk serie så att $p_t = \rho^t p_0 + c\frac{\rho^t - 1}{\rho - 1}$. Vi ser att serien konvergerar om och endast om $|\rho| = |-b/a| < 1$, dvs. om och endast om $b < a$.

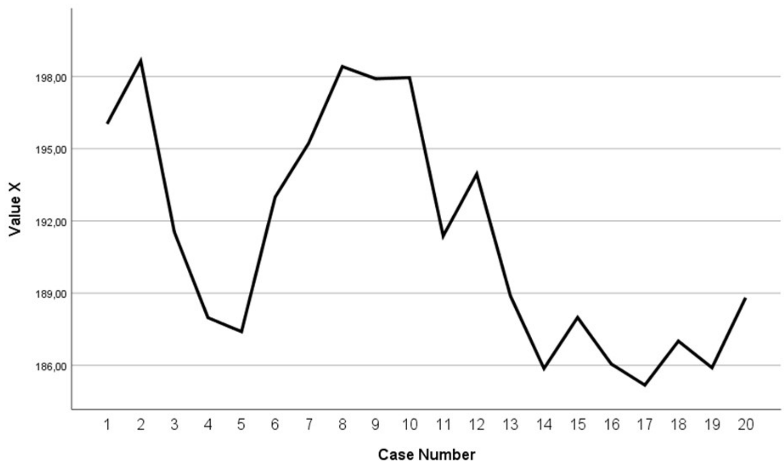
Om priset p_t konvergerar mot något tal, säg p^*, måste det gälla att $p^* = \rho p^* + c$. Löser vi ut p^* får vi $p^* = \frac{c}{1-\rho} = \frac{(d_0 - s_0)/a}{1-(-b/a)} = \frac{d_0 - s_0}{a+b}$.

För $a = 2$, $b = 1.6$, $d_0 = 160$, $s_0 = 120$ och $p_0 = 15$ får vi prisutvecklingen som visas i nedanstående figur. Vi ser hur priset varierar upp och ner, men stabiliserar sig kring $p^* = \frac{160-120}{2+1,6} = 11,11$

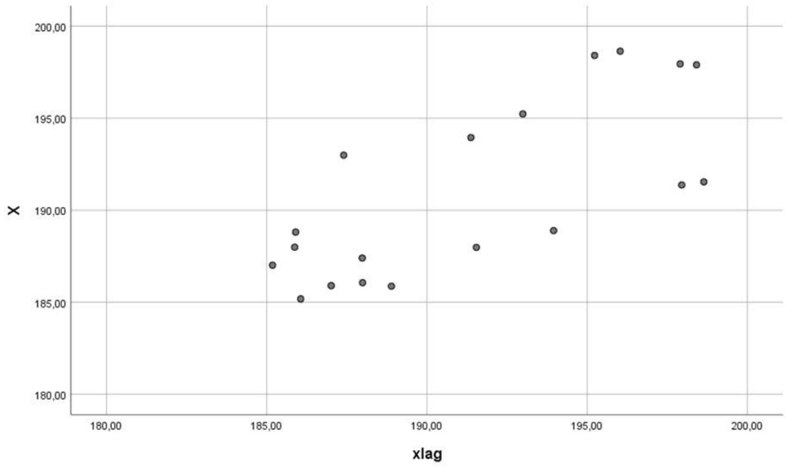


30.5

Tidsserien avser slutkurser på Googleaktien. En plot av tidsserien är



Vi önskar anpassa en första ordningens autoregressiv modell och plottar först x_t mot x_{t-1}



Figuren antyder ett linjärt samband mellan x_t och x_{t-1}. Vi beräknar nu erforderliga summor för uppskattning av modellparametrar

$$\sum x_t = 3629,07 \qquad \sum x_t^2 = 693572,83 \qquad n = 19$$
$$\sum x_{t-1} = 3636,29 \quad \sum x_{t-1}^2 = 696351,37 \quad \sum x_t x_{t-1} = 694850,30$$

$b = \frac{694850.30 - \frac{3629.07 \cdot 3636.29}{19}}{696351.37 - \frac{3636.29^2}{19}} = 0,7192$

$a = \frac{3629,07}{19} - 0,7192 \cdot \frac{3636,29}{19} = 53,361$

$SS_E = 187,356,\ s^2 = \frac{187,356}{19-2} = 11,021$

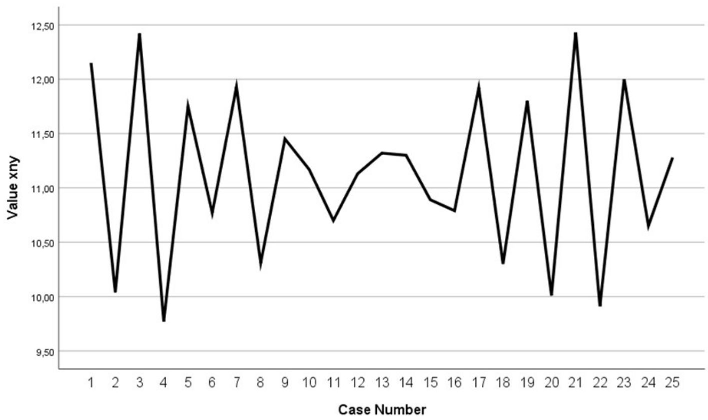
För att pröva $H_0 : \beta = 0$ beräknar vi $t_{obs} = \frac{0,7192}{\sqrt{11,021}/\sqrt{696351,37 - \frac{3636,29^2}{19}}} = 4,4651$. Eftersom $|t_{obs}| > t_{0,975}(17) = 2,11$ förkastar vi H_0 på 5 procentsnivån.

Alternativt kan vi pröva H_0 genom att beräkna $SS_T = 693572,83 - \frac{3629,07^2}{19} = 407,09$ så att $F_{obs} = \frac{407,09 - 187,356}{187,356/(19-2)} = 19,938$. Vi ser att $F_{obs} > F_{0,95}(1,17) = 4,45$ så att H_0 kan förkastas på 5 procentsnivån.

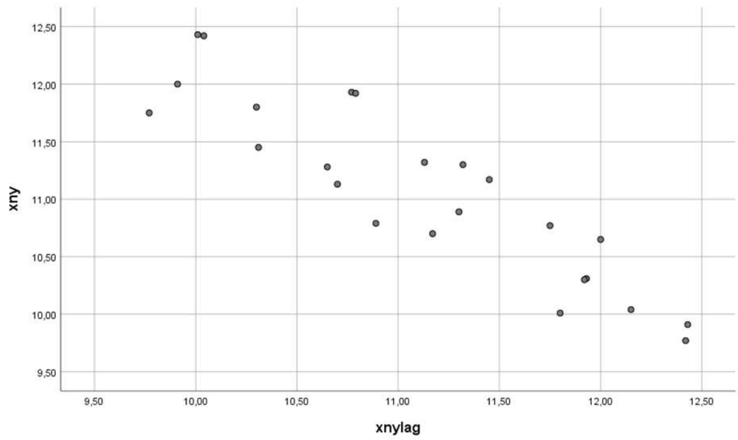
Prognos för X_{T+1} och X_{T+2} blir $53,361 + 0,7192 \cdot 188,81 = 189,15$ respektive $53,361 + 0,7192 \cdot 189,15 = 189,40$.

30.6

Låt X_t vara priset på bär i kr/kg vid tidpunkten t. Observationer på X_t visas i nedanstående figur



Vi antar att den betingade fördelningen för X_t givet att $X_{t-1} = x_{t-1}$ har väntevärdet $\mu_t = \alpha + \beta x_{t-1}$. Nedan visas en plot av x_t mot x_{t-1}



Figuren antyder att det finnas ett linjärt, negativt, samband mellan x_t mot x_{t-1} vilket motiverar modellen. Vi vill nu uppskatta parametrarna α och β

med minstakvadratmetoden. Följande summor erhålls:

$$\sum x_t = 266,04 \qquad \sum x_t^2 = 2963,26 \qquad n = 24$$
$$\sum x_{t-1} = 266,91 \quad \sum x_{t-1}^2 = 2983,65 \quad \sum x_t x_{t-1} = 2945,39$$

Detta ger

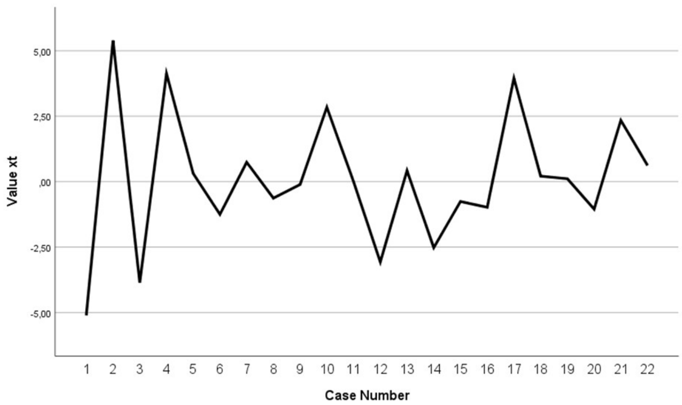
$b = \frac{2945{,}39 - \frac{266{,}04 \cdot 266{,}91}{24}}{2983{,}65 - \frac{266{,}91^2}{24}} = -0,87106$

$a = \frac{266{,}04}{24} - (-0,87106) \cdot \frac{266{,}91}{24} = 20,772$

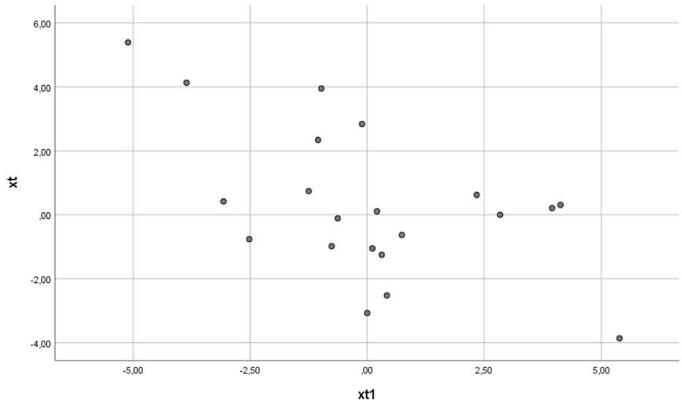
Processen konvergerar mot $\gamma = \frac{\alpha}{1-\beta}$ vilket uppskattas med $c = \frac{20{,}772}{1-(-0{,}87106)} = 11,102$ kr/kg.

30.7

Låt X_t vara dagliga relativa kursförändringar vid Stockholmsbörsen. En plot av X_t visas i nedanstående figur.



Vi antar att den betingade fördelningen för X_t givet att $X_{t-1} = x_{t-1}$ har väntevärdet $\mu_t = \alpha + \beta x_{t-1}$. Nedan visas en plot av x_t mot x_{t-1}



Figuren antyder inte något starkt samband mellan x_t mot x_{t-1}. För att uppskatta parametrarna α och β beräknar vi

$$\begin{array}{lll} \sum x_t = 6,83 & \sum x_t^2 = 111,80 & n = 21 \\ \sum x_{t-1} = 1,10 & \sum x_{t-1}^2 = 137,53 & \sum x_t x_{t-1} = -68,86 \end{array}$$

Detta ger

$$b = \frac{-68.86 - \frac{6.83 \cdot 1.10}{21}}{137.53 - \frac{1.10^2}{21}} = -0,5035$$

$$a = \frac{6.83}{21} - (-0.5035) \cdot \frac{1.10}{21} = 0,3516$$

Residualkvadratsumman beräknas till $SS_E = 74,73$ så att residualvariansen blir $s_e^2 = 74,72/19 = 3,9326$. En uppskattning av variansen för b blir därmed $s_b^2 = 3.9326/\left(137.53 - 1.10^2/21\right) = 0,02860$. För att pröva en hypotes om $\beta = 0$ får det observerade t-värdet $0.5035/\sqrt{0.02860} = 2,9773$, dvs. β är signifikant skild från 0 med signifikansnivån $0,05$.

31. χ^2-METODEN

31.1

H_0 : Kunderna föredrar alla märken lika mycket

H_A : Kunderna föredrar något märke bättre än övriga märken

Signifikansnivå $\alpha = 0,05$

χ^2-goodness of fit test

Antal frihetsgrader = 4 grupper - 1 = 3

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(3) = 7,81$

Observerade frekvenser:

Märke	A	B	C	D	Summa
Frekvens	28	41	81	50	200

Förväntade frekvenser:

Märke	A	B	C	D	Summa
Frekvens	50	50	50	50	200

$\chi^2_{obs} = \frac{(28-50)^2}{50} + \frac{(41-50)^2}{50} + \frac{(81-50)^2}{50} + \frac{(50-50)^2}{50} = 9,68 + 1,62 + 19,22 + 0,00 = 30,52$

H_0 förkastas på 5 %-nivån, dvs. data antyder att kunderna har större preferenser för något av de testade märkena.

31.2

H_0 : Observationerna stöder den genetiska teorin

H_A : Observationerna stöder inte den genetiska teorin

Signifikansnivå $\alpha = 0,05$

χ^2-goodness of fit test

Antal frihetsgrader = 4 grupper - 1 = 3

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(3) = 7,81$

a)

Observerade frekvenser:

Variant	gul/normal	gul/kort	svart/normal	svart/kort	Summa
Frekvens	57	20	12	11	100

Förväntade frekvenser:

Variant	gul/normal	gul/kort	svart/normal	svart/kort	Summa
Frekvens	$\frac{9}{16}100 = 56,25$	$\frac{3}{16}100 = 18,75$	$\frac{3}{16}100 = 18,75$	$\frac{1}{16}100 = 6,25$	100

$\chi^2_{obs} = \frac{(57-56,25)^2}{56,25} + \frac{(20-18,75)^2}{18,75} + \frac{(12-18,75)^2}{18,75} + \frac{(11-6,25)^2}{6,25} = 0,01 + 0,08 + 2,43 + 3,61 = 6,13$

H_0 kan inte förkastas på 5 %-nivån, dvs. det finns inget empiriskt stöd för att förkasta den genetiska teorin.

b)

Observerade frekvenser:

Variant	gul/normal	gul/kort	svart/normal	svart/kort	Summa
Frekvens	114	40	24	22	200

Förväntade frekvenser:

Variant	gul/normal	gul/kort	svart/normal	svart/kort	Summa
Frekvens	$\frac{9}{16}200 = 112,5$	$\frac{3}{16}200 = 37,5$	$\frac{3}{16}200 = 37,5$	$\frac{1}{16}200 = 12,5$	200

$\chi^2_{obs} = \frac{(57-56,25)^2}{56,25} + \frac{(20-18,75)^2}{18,75} + \frac{(12-18,75)^2}{18,75} + \frac{(11-6,25)^2}{6,25} = 0,02 + 0,17 + 4,86 + 7,22 = 12,27$

H_0 kan förkastas på 5 %-nivån, dvs. det finns empiriskt stöd för att förkaste den genetiska teorin.

31.3

Teorin säger att 50 % åker bil till arbetet, $\frac{2}{3}$ 50 % = 33,3 % åker kollektivtrafiken och $\frac{1}{3}$ 50 % = 16,7 % cyklar eller går till arbetet.

H_0 : De anställdas resor till arbetet följer teorin

H_A : De anställdas resor till arbetet följer inte teorin

Signifikansnivå $\alpha = 0,05$

χ^2-goodness of fit test

Antal frihetsgrader = 3 grupper - 1 = 2

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(2) = 5,99$

Observerade frekvenser:

Färdmedel	Bil	Kollektivt	Cykel/gå	Summa
Frekvens	22	10	18	50

Förväntade frekvenser:

Färdmedel	Bil	Kollektiv	Cykel/gå	Summa
Frekvens	25	16,67	8,33	50

$\chi^2_{obs} = \frac{(22-25)^2}{25} + \frac{(10-16{,}67)^2}{16{,}67} + \frac{(18-8{,}33)^2}{8{,}33} = 0,36 + 2,67 + 11,23 = 14,26$

H_0 förkastas på 5 %-nivån, dvs. data antyder att de anställda väljer färdmedel på ett annat sätt än den angivna teorin.

31.4

Teorin säger att längden av kronbladen är $N(\mu = 12, \sigma = 1)$. Ett blad klassificeras som litet om dess längd är mindre än 11 mm. Sannolikheten att ett slumpmässigt valt kronblad är litet är enligt teorin $P\left(Z < \frac{11-12}{1}\right) = \Phi(-1) = 0,1587$, där $Z \sim N(0,1)$. Ett blad klassificeras som stort om dess längd överstiger 13 mm. Sannolikheten att ett slumpmässigt valt blad är stort är enligt teorin $P\left(Z > \frac{13-12}{1}\right) = 1 - \Phi(1) = 0,1587$. Sannolikheten att ett slumpmässigt valt blad är mellanstort är därmed $1 - 0,1587 - 0,1587 = 0,682\,6$.

H_0 : Kronbladens storlek följer teorin

H_A : Kronbladens storlek följer inte teorin

Signifikansnivå $\alpha = 0,05$

χ^2-goodness of fit test

Antal frihetsgrader = 3 grupper - 1 = 2

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(2) = 5,99$

Observerade frekvenser:

Storlek	Litet	Mellan	Stort	Summa
Frekvens	5	35	10	50

Förväntade frekvenser:

Färdmedel	Litet	Mellan	Stort	Summa
Frekvens	0,1587·50=7,935	0,6826·50=34,13	0,1587·50=7,935	50

$\chi^2_{obs} = \frac{(5-7{,}935)^2}{7{,}935} + \frac{(35-34{,}13)^2}{34{,}13} + \frac{(10-7{,}935)^2}{7{,}935} = 1,09 + 0,02 + 0,54 = 1,65$

H_0 kan inte förkastas på 5 %-nivån, dvs. data ger inget stöd för att teorin är fel.

31.5

H_0 : Sjukfrånvaron är lika fördelad över veckans dagar

H_A : Sjukfrånvaron är inte lika fördelad över veckans dagar

Signifikansnivå $\alpha = 0,01$

χ^2-goodness of fit test

Antal frihetsgrader = 5 grupper - 1 = 4

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,99}(4) = 13,3$

Observerade frekvenser:

Veckodag	Måndag	Tisdag	Onsdag	Torsdag	Fredag	Summa
Frekvens	41	26	15	20	38	140

Förväntade frekvenser:

Veckodag	Måndag	Tisdag	Onsdag	Torsdag	Fredag	Summa
Frekvens	28	28	28	28	28	140

$\chi^2_{obs} = \frac{(41-28)^2}{28} + \frac{(26-28)^2}{28} + \frac{(15-28)^2}{28} + \frac{(20-28)^2}{28} + \frac{(38-28)^2}{28} = 6,04 + 0,14 + 6,04 + 2,29 + 3,57 = 18,07$

H_0 förkastas på 1 %-nivån, dvs. data ger stöd för att sjukfrånvaron är olika för olika veckodagar.

31.6

H_0 : Plantorna ger rosa, vita och röda blommor i proportionerna 2:1:1

H_A : Fördelningen av färger är inte i proportionerna 2:1:1

Signifikansnivå $\alpha = 0,05$

χ^2-goodness of fit test

Antal frihetsgrader = 3 grupper - 1 = 2

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(2) = 5,99$

Observerade frekvenser:

Färg	Rosa	Vit	Röd	Summa
Frekvens	75	20	25	120

Förväntade frekvenser:

Det finns sammanlagt 2+1+1=4 delar. Av 120 buskar förväntas därför $\frac{2}{4}120 = 60$ buskar ge rosa blommor, $\frac{1}{4}120 = 30$ ge vita blommor osv.

Märke	Rosa	Vit	Röd	Summa
Frekvens	60	30	30	120

$\chi^2_{obs} = \frac{(75-60)^2}{60} + \frac{(20-30)^2}{30} + \frac{(25-30)^2}{30} = 3,75 + 3,33 + 0,83 = 7,92$

H_0 förkastas på 5 %-nivån, dvs. data antyder att färgerna fördelar sig på något annat sätt än i proportionerna 2:1:1

31.7

H_0 : Efterfrågan av tidskriften följer sannolikhetsmodellen.

H_A : Efterfrågan av tidskriften följer inte sannolikhetsmodellen.

Signifikansnivå $\alpha = 0,05$

χ^2-goodness of fit test

Observerade frekvenser:

Efterfrågan	0	1	2	3 eller fler	Summa
Frekvens	32	12	6	0	50

Förväntade frekvenser:

Sannolikheten för 0 efterfrågade tidskrifter är $f(0) = 0,6\frac{0,5^0}{0!} = 0,6$

Förväntad frekvens med 0 efterfrågade tidskrifter vid 50 veckor är 0,6·50=30

Sannolikheten för 1 efterfrågad tidskrift är $f(1) = 0,6\frac{0,5^1}{1!} = 0,3$

Förväntad frekvens med 1 efterfrågad tidskrift vid 50 veckor är 0,3·50=15

Sannolikheten för 2 efterfrågade tidskrifter är $f(2) = 0,6\frac{0,5^2}{2!} = 0,075$

Förväntad frekvens med 2 efterfrågade tidskrifter vid 50 veckor är 0,075·50=3,75

Sannolikheten för 3 eller fler efterfrågade tidskrifter är $1-0,6-0,3-0,075 = 0,025$

Förväntad frekvens för 3 eller fler efterfrågade tidskrifter vid 50 veckor är 0,025·50=1,25

Förväntade frekvenser för utfallen 2 respektive 3 eller fler är mindre än 5 varför dessa grupper slås samman.

Efterfrågan	0	1	2 eller fler	Summa
Frekvens	30	15	5	50

Antal frihetsgrader = 3 grupper - 1 = 2

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(2) = 5,99$

$\chi^2_{obs} = \frac{(32-30)^2}{30} + \frac{(12-15)^2}{15} + \frac{(6-5)^2}{5} = 0,13 + 0,60 + 0,20 = 0,93$

H_0 förkastas inte på 5 %-nivån, dvs. data antyder att efterfrågan kan följa den föreslagna sannolikhetsmodellen.

31.8

H_0 : Antal nya objekt följer den angivna frekvensfunktionen

H_A : Antal nya objekt följer inte den angivna frekvensfunktionen

Signifikansnivå $\alpha = 0,05$

χ^2-goodness of fit test

Antal nya objekt	Antal veckor	Förväntat antal veckor
0	15	18,395
1	16	18,395
2	10	9,1975
3	6	3,0658
4 eller fler	3	0,9467
Summa	50	50,000

Förväntat antal veckor med 0 nya objekt fås genom $f(0) \cdot 50 = \frac{0,3679}{0!} \cdot 50 = 18,395$

Observera att klasserna ”3” och ”4 eller fler” har för få förväntade veckor (<5) så att de slås ihop med klassen ”2”. Då återstår 3 klasser.

Antal frihetsgrader = 3 klasser - 1 = 2

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(2) = 5,99$

$\chi^2_{obs} = \frac{(15-18.395)^2}{18,395} + \frac{(16-18,395)^2}{18,395} + \frac{(19-13,21)^2}{13,21} = 0,63 + 0,31 + 2,54 = 3,48$

H_0 förkastas inte på 5 %-nivån, dvs. data kan inte utesluta att antal nya objekt följer den föreslagna frekvensfunktionen

31.9

H_0 : Fördelningen av användningen av rabattkuponger är lika för de tre butikerna

H_A : Minst en butik har en annan fördelning av användningen av rabattkuponger är de övriga butikerna

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Antal frihetsgrader = (3 klasser - 1)(3 butiker - 1) = 4

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(4) = 9,49$

Observerade frekvenser:

Använder rabattkuponger	Butik A	Butik B	Butik C	Summa
Ofta	65	97	55	217
Ibland	98	89	77	264
Sällan	37	14	68	119
Summa	200	200	200	600

Förväntade frekvenser:

Använder rabattkuponger	Butik A	Butik B	Butik C	Summa
Ofta	72,33	72,33	72,33	217
Ibland	88,00	88,00	88,00	264
Sällan	39,67	39,67	39,67	119
Summa	200	200	200	600

$\chi^2_{obs} = \frac{(65-72{,}33)^2}{72{,}33} + \frac{(97-72{,}33)^2}{72{,}33} + \frac{(55-72{,}33)^2}{72{,}33} + \frac{(98-88)^2}{88} + \frac{(89-88)^2}{88} + \frac{(77-88)^2}{88} + \frac{(37-39{,}67)^2}{39{,}67} + \frac{(14-39{,}67)^2}{39{,}67} + \frac{(68-39{,}67)^2}{39{,}67}$

$= 0,7428 + 8,4143 + 4,1522 + 1,1364 + 0,0114 + 1,375 + 0,1797 + 16,6108 + 20,2316 = 52,8542$

H_0 förkastas på 5 %-nivån, dvs. ger ett stöd för att fördelningen för hur rabattkuponger används är olika för olika butiker

31.10

H_0 : Politikern är lika populär i alla delar av valdistriktet

H_A : Politikern är inte lika populär i alla delar av valdistriktet

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Antal frihetsgrader = (4 klasser - 1)(2 möjliga svar - 1) = 3

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(3) = 7,81$

Observerade frekvenser:

	Östra	Norra	Västra	Södra	Summa
Sympatisörer	21	29	23	16	89
Ej sympatisörer	39	41	27	24	131
Summa	60	70	50	40	220

Förväntade frekvenser:

	Östra	Norra	Västra	Södra	Summa
Sympatisörer	24,27	28,32	20,23	16,18	89
Ej sympatisörer	35,73	41,68	29,77	23,82	131
Summa	60	70	50	40	220

$\chi^2_{obs} = \frac{(21-24,27)^2}{24,27} + \frac{(29-28,32)^2}{28,32} + \frac{(23-20,33)^2}{20,33} + \frac{(16-16,18)^2}{16,18} + \frac{(39-35,73)^2}{35,73} + \frac{(41-41,68)^2}{41,68} + \frac{(27-29,77)^2}{29,77} + \frac{(24-23,82)^2}{23,82}$

$= 0,44 + 0,02 + 0,38 + 0,00 + 0,30 + 0,01 + 0,26 + 0,00 = 1,41$

H_0 förkastas inte på 5 %-nivån, dvs. data tyder inte på att fördelningen av sympatisörer är olika i olika delar av valdistriktet.

31.11

H_0 : Fördelningen av politiska sympatier är lika vid båda mättillfällena

H_A : Fördelningen av politiska sympatier är inte lika vid båda mättillfällena

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Observerade frekvenser:

	jan 2015	jan 2016	Summa
S	30	34	64
V	7	12	19
MP	7	9	16
M	25	40	65
C	6	10	16
L	5	9	14
KD	4	6	10
SD	13	27	40
FI	3	3	6
Summa	100	150	250

Förväntade frekvenser:

	jan 2015	jan 2016	Summa
S	25,6	38,4	64
V	7,6	11,4	19
MP	6,4	9,6	16
M	26	39	65
C	6,4	9,6	16
L	5,6	8,4	14
KD	4	6	10
SD	16	24	40
FI	2,4	3,6	6
Summa	100	150	250

Förväntade frekvenser är för få för KD och FI. Dessa förs därför samman till en klass med observerade frekvenser 7 respektive 9 och förväntade frekvenser 6,4 respektive 9,6.

Antal frihetsgrader = (8 klasser - 1)(2 fördelningar - 1) = 7

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0{,}95}(7) = 14,1$

$\chi^2_{obs} = \frac{(30-26{,}6)^2}{26{,}6} + \frac{(34-38{,}4)^2}{38{,}4} + \frac{(7-7{,}6)^2}{7{,}6} + \frac{(12-11{,}4)^2}{11{,}4} + \frac{(7-6{,}4)^2}{6{,}4} + \frac{(9-9{,}6)^2}{9{,}6} + \frac{(25-26)^2}{26} + \frac{(40-39)^2}{39} + \frac{(6-6{,}4)^2}{6{,}4} + \frac{(10-9{,}6)^2}{9{,}6} + \frac{(5-5{,}6)^2}{5{,}6} + \frac{(9-8{,}4)^2}{8{,}4} + \frac{(13-16)^2}{16} + \frac{(27-24)^2}{24} + \frac{(7-6{,}6)^2}{6{,}4} + \frac{(9-9{,}6)^2}{9{,}6}$

$= 0,76 + 0,50 + 0,05 + 0,03 + 0,06 + 0,04 + 0,04 + 0,03 + 0,03 + 0,02 + 0,06 + 0,04 + 0,56 + 0,38 + 0,02 + 0,04 = 2,66$

H_0 förkastas inte på 5 %-nivån, dvs. data tyder inte på att fördelningen av sympatisörer är olika vid de två tidpunkterna.

31.12

H_0 : Fördelningen av rökvanor är lika för män och kvinnor

H_A : Fördelningen av rökvanor är inte lika för män och kvinnor

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Observerade frekvenser:

	Kvinnor	Män	Summa
Röker ej	121	123	244
1-5 cigg/dag	18	3	21
6-15 cigg/dag	10	17	27
mer än 15 cigg/dag	3	5	8
Summa	152	148	300

Förväntade frekvenser:

	Kvinnor	Män	Summa
Röker ej	123,63	120,37	244
1-5 cigg/dag	10,64	10,36	21
6-15 cigg/dag	13,68	13,32	27
mer än 15 cigg/dag	4,05	3,95	8
Summa	152,0	148,0	300

Förväntade frekvenser är för få för klassen ”mer än 15 cigg/dag” och förs därför samman med klassen ”6 - 15 cigg/dag”.

Antal frihetsgrader = (3 klasser - 1)(2 fördelningar - 1) = 2

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(2) = 5,99$

$$\chi^2_{obs} = \frac{(121-123{,}63)^2}{123{,}63} + \frac{(123-120{,}37)^2}{120{,}37} + \frac{(18-10{,}64)^2}{10{,}64} + \frac{(3-10{,}36)^2}{10{,}36} + \frac{(13-17{,}73)^2}{14{,}73} + \frac{(22-17{,}27)^2}{17{,}27}$$

$$= 0,06 + 0,06 + 5,09 + 5,23 + 1,26 + 1,30 = 12,99$$

H_0 förkastas på 5 %-nivån, dvs. data tyder inte på att fördelningen av rökvanor är olika för kvinnor och män.

31.13

H_0 : Män och kvinnor ger lika betygsfördelning

H_A : Män och kvinnor ger olika betygsfördelning

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Observerade frekvenser:

	Män	Kvinnor	Summa
Dålig	6	11	17
Mellanbra	14	27	41
Bra	10	22	32
Summa	30	60	90

Förväntade frekvenser:

	Män	Kvinnor	Summa
Dålig	5,67	11,33	17
Mellanbra	13,67	27,33	41
Bra	10,67	21,33	32
Summa	30	60	90

Antal frihetsgrader = (3 klasser - 1)(2 fördelningar - 1) = 2

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(2) = 5,99$

$$\chi^2_{obs} = \frac{(6-5{,}67)^2}{5{,}67} + \frac{(11-11{,}33)^2}{11{,}33} + \frac{(14-13{,}67)^2}{13{,}67} + \frac{(27-27{,}33)^2}{27{,}33} + \frac{(10-10{,}67)^2}{10{,}67} + \frac{(22-21{,}33)^2}{21{,}33}$$

$= 0,02 + 0,01 + 0,01 + 0,00 + 0,04 + 0,02 = 0,10$

H_0 förkastas inte på 5 %-nivån, dvs. data tyder inte på att betygsfördelningen är olika för kvinnor och män.

31.14

H_0 : Stockholm och Göteborg har samma fördelning av politiska sympatier

H_A : Stockholm och Göteborg har inte samma fördelning av politiska sympatier

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Observerade frekvenser:

	Stockholm	Göteborg	Summa
M	184	94	278
C	8	6	14
FP	39	22	61
KD	12	3	15
MP	71	40	111
S	104	67	171
V	20	29	49
SD	5	4	9
Övriga	4	4	8
Summa	447	269	716

Förväntade frekvenser:

	Stockholm	Göteborg	Summa
M	173,56	104,44	278
C	8,74	5,26	14
FP	38,08	22,92	61
KD	9,36	5,64	15
MP	69,30	41,70	111
S	106,76	64,24	171
V	30,59	18,41	49
SD	5,62	3,38	9
Övriga	4,99	3,01	8
Summa	447	269	716

Förväntade frekvenser är för få för SD och Övriga. Dessa förs därför samman till en klass med observerade frekvenser 9 respektive 8 och förväntade frekvenser 10,61 respektive 6,39.

Antal frihetsgrader = (8 klasser - 1)(2 fördelningar - 1) = 7

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0{,}95}(7) = 14,1$

$$\chi^2_{obs} = \frac{(184-173{,}56)^2}{173{,}56} + \frac{(94-104{,}44)^2}{104{,}44} + \frac{(8-8{,}74)^2}{8{,}74} + \frac{(6-5{,}26)^2}{5{,}26} + \frac{(39-38{,}08)^2}{38{,}08} + \frac{(22-22{,}92)^2}{22{,}92} + \frac{(12-9{,}36)^2}{9{,}36} + \frac{(3-5{,}64)^2}{5{,}64} + \frac{(71-69{,}30)^2}{69{,}30} + \frac{(40-41{,}70)^2}{41{,}70} + \frac{(104-106{,}76)^2}{106{,}76} + \frac{(67-64{,}24)^2}{64{,}24} + \frac{(20-30{,}59)^2}{30{,}59} + \frac{(29-18{,}41)^2}{18{,}41} + \frac{(9-10{,}61)^2}{10{,}61} + \frac{(8-6{,}39)^2}{6{,}39}$$

$$= 0,63 + 1,04 + 0,06 + 0,10 + 0,02 + 0,04 + 0,74 + 1,23 + 0,04 + 0,07 + 0,07 + 0,12 + 3,67 + 6,09 + 0,25 + 0,41 = 14,59$$

H_0 förkastas på 5 %-nivån, dvs. det finns empiriskt evidens för att fördelningen av politiska sympatier är olika i Stockholm och Göteborg.

31.15

a)

H_0 : Fördelningen av resvanor till och från arbetet är lika för män och kvinnor

H_A : Fördelningen av resvanor till och från arbetet är lika för män och kvinnor

b)

Observationer på resvanor på de 50 männen och 50 kvinnorna är oberoende av varandra. Förväntade frekvenser ska vara större än 5.

c)

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Observerade frekvenser:

	Män	Kvinnor	Summa
Egen bil	22	6	28
Buss	4	21	25
Cykel	16	11	27
Annat	8	12	20
Summa	50	50	100

Förväntade frekvenser:

	Män	Kvinnor	Summa
Egen bil	14	14	28
Buss	12,5	12,5	25
Cykel	13,5	13,5	27
Annat	10	10	20
Summa	50	50	100

Förväntade frekvenser är tillräckligt stora i alla klasser.

Antal frihetsgrader = (4 klasser - 1)(2 fördelningar - 1) = 3

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(3) = 7,81$

$\chi^2_{obs} = \frac{(22-14)^2}{14} + \frac{(6-14)^2}{14} + \frac{(4-12{,}5)^2}{12{,}5} + \frac{(21-12{,}5)^2}{12{,}5} + \frac{(16-13{,}5)^2}{13{,}5} + \frac{(11-13{,}5)^2}{13{,}5} + \frac{(8-10)^2}{10} + \frac{(12-10)^2}{10}$

$= 4,57 + 4,57 + 5,78 + 5,78 + 0,46 + 0,46 + 0,40 + 0,40 = 22,43$

H_0 förkastas på 5 %-nivån, dvs. det finns empiriskt evidens för att män och kvinnor vid företaget har olika vanor vad beträffar resor till och från arbetet.

31.16

H_0 : Fördelningen av fiskarter är lika olika månader

H_A : Fördelningen av fiskarter är inte lika olika månader

Signifikansnivå $\alpha = 0,05$

χ^2-homogenitetstest

Observerade frekvenser:

	Juni	Juli	Augusti	September	Summa
Röding	31	27	29	24	111
Forell	134	125	148	120	527
Harr	62	51	71	49	233
Summa	227	203	248	193	871

Förväntade frekvenser:

	Juni	Juli	Augusti	September	Summa
Röding	28,93	25,87	31,61	24,60	111
Forell	137,35	122,83	150,05	116,77	527
Harr	60,72	54,30	66,34	51,63	233
Summa	227	203	248	193	871

Förväntade frekvenser är tillräckligt stora i alla klasser.

Antal frihetsgrader = (3 klasser - 1)(4 fördelningar - 1) = 6

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(6) = 12,6$

$\chi^2_{obs} = \frac{(31-28{,}93)^2}{28{,}93} + \frac{(27-25{,}87)^2}{25{,}87} + \frac{(29-31{,}61)^2}{31{,}61} + \frac{(24-24{,}60)^2}{24{,}60} + \frac{(134-137{,}35)^2}{137{,}35} + \frac{(125-122{,}83)^2}{122{,}83} + \frac{(148-150{,}05)^2}{150{,}05} + \frac{(120-116{,}77)^2}{116{,}77} + \frac{(62-60{,}72)^2}{60{,}72} + \frac{(51-54{,}30)^2}{54{,}30} + \frac{(71-66{,}34)^2}{66{,}34} + \frac{(49-51{,}63)^2}{51{,}63}$

$= 0,15 + 0,05 + 0,21 + 0,01 + 0,08 + 0,04 + 0,03 + 0,09 + 0,03 + 0,20 + 0,33 + 0,13 = 1,35$

H_0 förkastas inte på 5 %-nivån, dvs. det finns inget empiriskt stöd för att fördelningen av fiskarter är olika vid olika månader.

31.17

H_0 : Det finns inget beroende mellan utbildning och inställning till äktenskap

H_A : Det finns ett beroende mellan utbildning och inställning till äktenskap

Signifikansnivå $\alpha = 0,05$

Test av oberoende med χ^2-metoden

Observerade frekvenser:

Utbildning	Inställning till äktenskap				
	Mycket dålig	Dålig	God	Mycket god	Totalt
Grundskola	18	29	70	115	232
Gymnasium	17	28	30	41	116
Universitet	11	10	11	20	52
Totalt	46	67	111	176	400

Förväntade frekvenser:

Utbildning	Inställning till äktenskap				
	Mycket dålig	Dålig	God	Mycket god	Totalt
Grundskola	26,68	38,86	64,38	102,08	232
Gymnasium	13,34	19,43	32,19	51,04	116
Universitet	5,98	8,71	14,43	22,88	52
Totalt	46	67	111	176	400

Förväntade frekvenser är tillräckligt stora i alla celler.

Antal frihetsgrader = (3 rader - 1)(4 kolumner - 1) = 6

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(6) = 12,6$

$\chi^2_{obs} = \frac{(18-26,68)^2}{26,68} + \frac{(29-38,86)^2}{38,86} + \frac{(70-64,38)^2}{64,38} + \frac{(115-102,08)^2}{102,08} + \frac{(17-13,34)^2}{13,34} + \frac{(28-19,43)^2}{19,43} + \frac{(30-32,19)^2}{32,19} + \frac{(41-51,04)^2}{51,04} + \frac{(11-5,98)^2}{5,98} + \frac{(10-8,71)^2}{8,71} + \frac{(11-14,43)^2}{14,43} + \frac{(20-22,88)^2}{22,88}$

$= 2,82 + 2,50 + 0,49 + 1,64 + 1,00 + 3,78 + 0,15 + 1,97 + 4,21 + 0,19 + 0,82 + 0,36 = 19,94$

H_0 förkastas på 5 %-nivån, dvs. det finns empiriskt stöd för att inställning till äktenskap och utbildning är beroende.

31.18

H_0 : Uppfattning om rökning är skadligt är oberoende av om man röker eller ej

H_A : Uppfattning om rökning är skadligt är inte oberoende av om man röker eller ej

Signifikansnivå $\alpha = 0,05$

Test av oberoende med χ^2-metoden

Observerade frekvenser:

Rökare	Anser rökning är skadlig		
	Ja	Nej	Totalt
Ja	80	80	160
Nej	25	15	40
Totalt	105	95	200

Förväntade frekvenser:

Rökare	Anser rökning är skadlig		
	Ja	Nej	Totalt
Ja	84	76	160
Nej	21	19	40
Totalt	105	95	200

Förväntade frekvenser är tillräckligt stora i alla celler.

Antal frihetsgrader = (2 rader - 1)(2 kolumner - 1) = 1

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(1) = 3,84$

$\chi^2_{obs} = \frac{(80-84)^2}{84} + \frac{(80-76)^2}{76} + \frac{(25-21)^2}{21} + \frac{(15-19)^2}{19}$

$= 0,19 + 0,21 + 0,76 + 84 = 2,01$

H_0 förkastas inte på 5 %-nivån, dvs. det finns inget empiriskt stöd för beroende mellan uppfattningen om rökning är skadlig och om man själv röker.

31.19

H_0 : Det finns inget beroende mellan stadsdel och teatervanor

H_A : Det finns ett beroende mellan stadsdel och teatervanor

Signifikansnivå $\alpha = 0,05$

Test av oberoende med χ^2-metoden

Observerade frekvenser:

Antal	Stadsdel				
teaterbesök	Södermalm	Kungsholmen	Norrmalm	Östermalm	Totalt
Inga	54	45	50	40	189
1-2	36	32	28	29	125
>2	27	18	16	25	86
Totalt	117	95	94	94	400

Förväntade frekvenser:

Antal	Stadsdel				
teaterbesök	Södermalm	Kungsholmen	Norrmalm	Östermalm	Totalt
Inga	55,28	44,89	44,42	44,42	189
1-2	36,56	29,69	29,38	29,38	125
>2	25,16	20,43	20,21	20,21	86
Totalt	117	95	94	94	400

Förväntade frekvenser är tillräckligt stora i alla celler.

Antal frihetsgrader = (3 rader - 1)(4 kolumner - 1) = 6

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(6) = 12,6$

$$\chi^2_{obs} = \frac{(54-55,28)^2}{55,28} + \frac{(45-44,89)^2}{44,89} + \frac{(50-44,42)^2}{44,42} + \frac{(40-44,42)^2}{44,42} + \frac{(36-36,56)^2}{36,56} + \frac{(32-29,69)^2}{29,69} + \frac{(28-29,38)^2}{29,38} + \frac{(29-29,38)^2}{29,38} + \frac{(27-25,16)^2}{25,16} + \frac{(18-20,43)^2}{20,43} + \frac{(16-20,21)^2}{20,21} + \frac{(25-20,21)^2}{20,21}$$

$= 0,03 + 0,00 + 0,70 + 0,44 + 0,01 + 0,18 + 0,06 + 0,00 + 0,14 + 0,29 + 0,88 + 1,14 = 3,86$

H_0 förkastas inte på 5 %-nivån, dvs. det finns inget empiriskt stöd för ett beroende mellan stadsdel och teatervanor.

31.20

H_0 : Det finns inget beroende mellan inkomst och konsumtion

H_A : Det finns ett beroende mellan inkomst och konsumtion

Signifikansnivå $\alpha = 0,05$

Test av oberoende med χ^2-metoden

Observerade frekvenser:

Konsumtion	Inkomst			
	Låg	Mellan	Hög	Totalt
Låg	29	25	3	57
Mellan	12	35	11	58
Hög	1	9	15	25
Totalt	42	69	29	140

Förväntade frekvenser:

Konsumtion	Inkomst			
	Låg	Mellan	Hög	Totalt
Låg	17,10	28,09	11,81	57
Mellan	17,40	28,59	12,01	58
Hög	7,50	12,32	5,18	25
Totalt	42	69	29	140

Förväntade frekvenser är tillräckligt stora i alla celler.

Antal frihetsgrader = (3 rader - 1)(3 kolumner - 1) = 4

Förkasta H_0 om $\chi^2_{obs} > \chi^2_{0,95}(4) = 9,49$

$$\chi^2_{obs} = \frac{(29-17{,}10)^2}{17{,}10} + \frac{(25-28{,}09)^2}{28{,}09} + \frac{(3-11{,}81)^2}{11{,}81} + \frac{(12-17{,}40)^2}{17{,}40} + \frac{(35-28{,}59)^2}{28{,}59} + \frac{(11-12{,}01)^2}{12{,}01} + \frac{(1-7{,}50)^2}{7{,}50} + \frac{(9-12{,}32)^2}{12{,}32} + \frac{(15-5{,}18)^2}{5{,}18}$$

$$= 8,28 + 0,34 + 6,57 + 1,68 + 1,44 + 0,09 + 5,63 + 0,90 + 18,63 = 43,55$$

H_0 förkastas på 5 %-nivån, dvs. det finns empiriskt stöd för att det finns ett beroende mellan inkomst och konsumtion.

32. BESLUTSTEORI

32.1

Förväntad vinst är $35 \cdot \frac{1}{37} + 0 \cdot \frac{36}{37} = \frac{35}{37}$. Insatsen är 1, dvs. förväntad vinst är mindre än insatsen varför man bör avstå från att spela.

32.2

Beslutsmatrisen är

Antal beställda tidningar	Efterfrågat antal tidningar 16	17	18	19	20	21	Förväntad vinst
Sannolikhet	0,05	0,12	0,16	0,18	0,36	0,13	
16	64	64	64	64	64	64	64,00
17	58	68	68	68	68	68	67,50
18	52	62	72	72	72	72	69,80
19	46	56	66	76	76	76	70,50
20	40	50	60	70	80	80	69,40
21	34	44	54	64	74	84	64,70

Försäljaren bör beställa 19 tidningar

32.3

Sannolikheten att vinsten blir 2 kr är $\frac{1}{2}$, sannolikheten att vinsten blir 4 kr är $\frac{1}{4}$ osv. så att sannolikheten att vinsten blir 2^n är $\frac{1}{2^n}$. Den förväntade vinsten blir därför

$$2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + \ldots + 2^n \frac{1}{2^n} + \ldots = 1 + 1 + \ldots + 1 + \ldots$$

dvs. den förväntade vinsten är oändligt stor. Eftersom förväntad vinst är större än insatsen (1000 kr) bör man delta i spelet.

För att nettot från spelet ska vara positivt måste vinsten överstiga insatsen, 1000 kr. Detta inträffar om vinsten är 1024 kr eller mer. Eftersom $2^{10} = 1024$ krävs det därför att det inträffar minst 9 krona innan klave inträffar för första gången. Om det blir 8 eller färre klave innan krona inträffar för första gången gör man en förlust. Detta inträffar med sannolikheten

$\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots + \frac{1}{2^9} = 0.998\,05$

dvs. sannolikheten att spelet slutar med förlust är mycket stor och man bör av den anledningen inte delta i spelet.

Detta är en paradox eftersom den förväntade vinsten är oändligt stor, men med nästan praktisk säkerhet kommer man att förlora.

32.4

Beslutsmatrisen är

Handlings alternativ	Naturtillstånd		Minsta nyttan	Största nyttan	Förväntad nytta enl. Laplace
	Färskt ägg	Surt ägg			
Använd ägget	5	0	0	5	2,5
Använd inte ägget	4	4	4	4	4
Maximal nytta	5	4			

Den minsta nyttan maximeras om man inte använder ägget, dvs. enligt maximinkriteriet ska vi inte använda det extra ägget

Den största nyttan maximeras om man väljer att använda ägget, dvs. enligt maximaxkriteriet ska vi använda det extra ägget

Regretmatris:

Handlings alternativ	Naturtillstånd		Maximal regret
	Ägget är färskt	Ägget är surt	
Använd ägget	0	4	4
Använd inte ägget	1	0	1

Maximal regret minimeras om man inte använder ägget, dvs. enligt minimax-regretkriteriet ska vi inte använda det extra ägget

Den förväntade nyttan enligt Laplacekriteriet maximeras om man inte använder ägget, dvs. enligt Laplacekriteriet ska vi inte använda det extra ägget

32.5

Beslutsmatrisen är

Antal beställda lådor	Efterfrågat antal lådor 25	26	27	28	29	30	min	max	Laplace
25	250	250	250	250	250	250	250	250	250,00
26	245	260	260	260	260	260	245	260	257,50
27	240	255	270	270	270	270	240	270	262,50
28	235	250	265	280	280	280	235	280	265,00
29	230	245	260	275	290	290	230	290	265,00
30	225	240	255	270	285	300	225	300	262,50
max	250	260	270	280	290	300			

Enligt maximinkriteriet ska 25 lådor beställas

Enligt maximaxkriteriet ska 30 lådor beställas

Enligt Laplacekriteriet ska 28 eller 29 lådor beställas

Regretmatris

Antal beställda lådor	Efterfrågat antal lådor 25	26	27	28	29	30	max
25	0	10	20	30	40	50	50
26	5	0	10	20	30	40	40
27	10	5	0	10	20	30	30
28	15	10	5	0	10	20	20
29	20	15	10	5	0	10	20
30	25	20	15	10	5	0	25

Enligt minimax regretkriteriet ska 28 eller 29 lådor beställas

32.6

a) Beslutsmatris:

Investerings alternativ	Naturtillstånd/marknadens utveckling			Förväntad vinst
	Stark utveckling	Moderat utveckling	Svag utveckling	
Sannolikhet	0,2	0,5	0,3	
Obligationer	12000	12000	12000	12000
Lågriskfond	43000	12000	-6000	12800
Högriskfond	66000	8000	-15000	12700

Förväntad vinst för obligationer: 0,2·12000+0,5·12000+0,3·12000=12000

Förväntad vinst för lågriskfond: 0,2·43000+0,5·12000+0,3·(-6000) = 12800

Förväntad vinst för obligationer: 0,2·66000+0,5·8000-0,3·15000 = 12700

Den förväntade vinsten maximeras om man väljer att investera i lågriskfonden.

32.7

Inköpt mängd glögg	Väder vid julmarknaden				Laplace
	Regnigt	Mulet	Solsken	min	
50 liter	2500	2500	2500	2500	2500,00
100 liter	0	5000	5000	0	3333,33
150 liter	-2500	2500	7500	-2500	2500,00
max	2500	5000	7500		

a) Enligt maximinkriteriet ska Nisse beställa 50 liter

Regretmatris

Inköpt mängd glögg	Väder vid julmarknaden			
	Regnigt	Mulet	Solsken	max
50 liter	0	2500	5000	5000
100 liter	2500	0	2500	2500
150 liter	5000	2500	0	5000
max	2500	5000	7500	

b) Enligt minimax regretkriteriet ska Nisse beställa 100 liter

c) Enligt Laplacekriteriet ska Nisse beställa 100 liter

32.8

Beslutsmatris:

Charlies handlingsalternativ	Pizza Queens alternativ Liten	Mellan	Stor	Minsta nyttan	Förv. nytta enl. Laplace
Liten	60	50	20	20	43,33
Mellan	50	60	10	10	40
Stor	90	60	0	0	50
Största nyttan	90	60	20		

a) Den minsta nyttan maximeras av en liten annons, dvs. enligt maximinkriterier ska Charlie välja en liten annons

b) Regretmatris

Charlies handlingsalternativ	Pizza Queens alternativ Liten	Mellan	Stor	Största regret
Liten	30	10	0	30
Mellan	40	0	10	40
Stor	0	0	20	20

Största regret minimeras av en stor annons, dvs. enligt minimaxregretkriteriet ska Charlie välja en stor annons

c) Förväntad nytta enligt Laplace kriteriet är för liten annons: $\frac{1}{3}60+\frac{1}{3}50+\frac{1}{3}20=43{,}33$, för mellanstor annons $\frac{1}{3}50+\frac{1}{3}60+\frac{1}{3}10=40$ och för stor annons $\frac{1}{3}90+\frac{1}{3}60+\frac{1}{3}0=50$, dvs. enligt Laplacekriteriet ska Charlie välja en stor annons.

32.9

a) Beslutsmatris

Försäkringsalternativ	Olycka	Stöld	Inget händer	min	Laplace
Olycksfall	-100	-8100	-100	-8100	-900
Stöld	-8150	-1150	-150	-8150	-330
Allrisk	-200	-1200	-200	-1200	-300
Ingen försäkring	-8000	-8000	0	-8000	-880

b) Möjlig kostnad minimeras om han tecknar en allriskförsäkring

c) Förväntad kostnad minimeras om han tecknar en allriskförsäkring

32.10

Beslutsmatris:

Beställda exemplar	Efterfrågan 10	11	12	13	14	Minsta nyttan
10	150	150	150	150	150	150
11	135	165	165	165	165	135
12	120	150	180	180	180	120
13	105	135	165	195	195	105
14	90	120	150	180	210	90

Minsta nyttan maximeras då Viktoria beställer 10 tidningar, dvs. enligt maximinkriteriet ska hon beställa hem 10 tidningar.