

# Tillämpad statistik

Uppgifter sammanställda av  
Eva Leander,  
Claudia Libiseller,  
Stig Danielsson,  
Karl Wahlin och  
Isak Hietala

Reviderad inför HT16

Institutionen för Datavetenskap  
Avdelningen för Statistik och Maskininlärning  
Linköpings universitet

Kompendium med extra  
övningsuppgifter

# Kapitel 2

---

## Uppgift 2:1

Man har i en kommun gjort en studie av kostnaderna för färdtjänst. I kommunen finns 3800 färdtjänstberättigade personer. Personerna delas in i sex grupper efter kön och var i kommunen som man bor. I tabellen nedan redovisas den genomsnittliga kostnaden per person för färdtjänst under en vecka ett visst år.

Genomsnittlig kostnad per person den aktuella veckan:

Boende	Kvinnor	Män
Centralt i tätorten	82 kr	80 kr
I utkanten av tätorten	130 kr	120 kr
På landet	162 kr	160 kr

Av tabellen nedan framgår hur de 3800 personerna fördelar sig på de sex grupperna.

Boende	Kvinnor	Män
Centralt i tätorten	1700	280
I utkanten av tätorten	500	220
På landet	300	800
Totalt	2500	1300

- Ange för var och en av de två tabellerna om de är av typen frekvens- eller medeltalstabell, respektive hur mångvägsindelade de är.
- Beskriv med hjälp av den andra tabellen fördelningen av boendes kön och område med såväl absoluta som relativa frekvenser. Detta kallas även marginalfördelning (se kapitel 9).
- Beskriv fördelningen av område enskilt för män respektive kvinnor med relativa frekvenser. Respektive kolumn ska alltså summera till 100 procent. Detta kallas även betingade fördelningar.

- d) Beräkna med hjälp av uppgifterna i de båda tabellerna den genomsnittliga kostnaden för färdtjänst för kvinnor respektive för män under den aktuella veckan.
- e) Beräkna standardvägda medeltal för färdtjänstkostnaden för kvinnor respektive män under den aktuella veckan. Använd det totala antalet boende i områdena som standardvikter.
- f) Jämför kostnaden för kvinnor och män med ledning av dels de båda medeltalen från punkt 4 och dels de båda standardvägda medeltalen från punkt 5. Försök att förklara varför de båda könsjämförelserna blir olika.

## Uppgift 2:2

I bokhandeln vid en högskola arbetar fyra heltidsanställda personer (som dock har flexitid) samt ett antal timanställda, som arbetar mellan 2 och 10 timmar per vecka under terminstid. Under en majvecka noterade man följande arbetstider för de 16 personer som arbetat den aktuella veckan:

42, 36, 39, 35, 8, 10, 8, 5, 7, 6, 2, 9, 8, 8, 9, 8

- a) Beräkna medelvärde och median i datamaterialet. Förklara varför de båda måtten skiljer sig åt på det sätt som de gör.
- b) Beräkna standardavvikelsen i datamaterialet.
- c) Beskriv datamaterialet i ett lådagram.
- d) Beräkna medelarbetstiden bland de 4 heltidsanställda (de fyra första personerna) respektive bland de 12 timanställda. Bilda sedan medelarbetstiden bland samtliga anställda som ett vägt medelvärde av de båda medelarbetstiderna.

### Uppgift 2:3

Sextio studenter registreras en viss termin på en fristående 20-poängskurs vid en högskola. Kursen är uppbyggd av fyra 5-poängsmoduler, som tenteras en och en. Efter en termin kan man därmed ha uppnått högst 20 poäng och minst 0 poäng på kursen. Vid terminens slut sammanfattas resultaten med avseende på antal avklarade poäng för de 60 studenterna enligt följande tabell:

Antal poäng	Antal personer
0	6
5	8
10	10
15	16
20	20

- Beräkna medelantalet avklarade poäng per registrerad student.
- Beräkna standardavvikelsen bland de 60 poängtal.
- Beskriv studieresultaten för de 60 studenterna i ett lådagram.

De ekonomiska resurser som institutionen disponerar för att ge den aktuella kursen baseras på två storheter, nämligen så kallade "Helårsstudenter" (HÅS) och "Helårsprestationer" (HÅP). En person som är registrerad på kurser omfattande 20 poäng under en termin motsvarar en halv HÅS medan en total prestation om 20 poäng under en termin motsvarar en halv HÅP. Antag att en hel HÅS utgör 7 tkr och att en hel HÅP utgör 12 tkr.

- Beräkna det belopp som institutionen kan disponera för att genomföra den aktuella kursen.

## Uppgift 2:4

På studentcentrum vid en mindre högskola har man efter antagningen en hösttermin sammanställt statistik över antal sökande uppdelat i antal antagna och antal som ej antagits för vardera könet och för de olika utbildningsprogrammen. Som exempel presenteras här data för två utbildningsprogram, ett ingenjörprogram med 330 platser och ett psykologprogram med 55 platser.

Utbildningsprogram	Kön	Får plats	Får ej plats
Ingenjörutbildning	Kvinnor	90	10
Ingenjörutbildning	Män	240	60
Psykologutbildning	Kvinnor	50	100
Psykologutbildning	Män	5	45

- Beräkna procentandelen som får plats bland de 250 kvinnorna respektive bland de 350 männen. Vilket av könen har högst andel antagna?
- Bilda en tvåvägsindeld tabell med två kolumner och två rader svarande mot de två könen (kolumner) och de två utbildningsprogrammen (rader). I varje ruta anges procentandelen antagna. Vilket av könen har högst andel antagna?
- Jämför resultaten från a) och b) och försök förklara varför det blir som det blir.

## Uppgift 2:5

De tolv patienterna vid ett serviceboende beskrivs med avseende på hur stort hjälpbehovet avseende städning, inköp och personlig omvårdnad bedömts vara i timmar per vecka och hur väl de boende trivs samt kön och ålder.

Person nr	Kön	Ålder	Hjälpbehov, tim/vecka	Trivsel i boendet
1	K	76	2	god
2	M	82	6	god
3	M	93	11	dålig
4	K	98	10	dålig
5	K	87	3	dålig
6	M	81	3	god
7	K	79	3	god
8	K	95	14	dålig
9	K	88	8	dålig
10	M	81	2	god
11	M	81	10	god
12	K	79	12	god

- Beskriv hur de boende fördelar sig på ålder i ett prickdiagram.
- Beskriv i ett lämpligt diagram hur de boende fördelar sig på hjälpbehov.
- Beskriv sambandet mellan kön och trivsel i boendet i en tvåvägsindeldad frekvenstabell.
- Trivs män eller kvinnor bäst? Kan skillnaden mellan könen snarare vara ett uttryck för skillnaden mellan åldrar? Utred.

## Uppgift 2:6

Vid en vårdcentral finns 16 personer anställda. Deras månadslöner för april ett visst år i tusental kronor ges, tillsammans med uppgift om ålder och kön enligt:

Person, nr	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Lön, tkr	15	17	17	31	26	20	24	16	20	18	26	30	30	26	34	22
Ålder	21	26	23	32	39	22	43	57	63	42	28	32	47	44	59	49
Kön	k	k	k	k	k	k	k	k	k	k	m	m	m	m	m	m

a) Beräkna medellönen för samtliga anställda i april.

b) Beräkna medellön

- för männen
- för kvinnorna
- för de yngre (under 40 år)
- för de äldre (40 år eller mer)

Beräkna också skillnaden i medellön mellan män och kvinnor respektive mellan äldre och yngre.

c) Bilda en tabell särskilt lämpad att jämföra de yngre och äldre med avseende på lön så att man också kan se skillnaden mellan könen. Lämpligen gör man då en fyrfältstabell med två kolumner, en för yngre och en för äldre, och två rader, en för kvinnor och en för män. I varje ruta placeras medelvärdet för den aktuella gruppen samt antalet personer.

d) Beräkna skillnaden i medellön mellan yngre och äldre bland männen och bland kvinnorna. Jämför dessa skillnader med skillnaden mellan åldersgrupperna i b).

e) Vilka tjänar egentligen genomsnittligt bäst, yngre eller äldre? Utred frågan.

f) Kan man hävda att det föreligger lönediskriminering vid vårdcentralen? Skulle slutsatsen förändras om du fick veta att personerna nr 4, 5, 11, 12, 13, 14 och 15 är läkare och/eller chefer medan de övriga är annan vårdpersonal och administrativ personal? Formulera några tankar.

## Uppgift 2:7

I en studie syftande till att kartlägga hur äldre utnyttjar sin tid, har man tagit reda på hur lång tid de undersökta personerna lagt på att läsa dagstidningar. Här följer data avseende 18 äldre kvinnor. Observationerna anger antalet minuter per dag som personerna uppgett sig läsa dagstidning en "vanlig" vardag, benämmt lästid.

5 5 10 10 14 40 60 80  
0 0 17 30 35 50  
5 15 18 20

Beräkna genomsnittlig lästid per person i gruppen om 18 kvinnor.

## Uppgift 2:8

Om 22 äldre mäns tidningsläsande får man veta den genomsnittliga lästiden för olika åldersgrupper, se tabellen som följer. Tabellen visar t ex att 5 män i åldern 70-79 år i genomsnitt haft en lästid om 25 minuter.

Ålder	Genomsnittlig lästid	Antal personer
60-69 år	35 min	5
70-79 år	25 min	5
80-89 år	17 min	12

a) Beräkna genomsnittlig lästid per person i gruppen om 22 män.

Betrakta nu kvinnornas lästider från föregående uppgift. Det gäller att de åtta tiderna på första raden (5 5 10 ...) avser kvinnor i åldern 60-69 år, de sex tiderna på nästa rad avser kvinnor i åldern 70-79 år och de fyra tiderna på den sista raden avser kvinnor i åldrarna 80-89 år.

b) Konstruera för de 18 kvinnorna en tabell liknande den som gavs för männen ovan. Jämför därefter de genomsnittliga lästiderna i de olika åldersgrupperna för män och kvinnor. Förklara i enkla ordalag hur jämförelsen mellan könen utfaller.

c) Jämför nu kvinnornas medelvärde från föregående uppgift med männens från a) i denna uppgift. Kommentera vad man ser och förklara skillnaden mot resultatet b). Vilketdera könet läser egentligen mest och hur uppstår paradoxen?

d) Bilda standardvägda medeltal för lästiden för männen respektive kvinnorna. Välj



standardvikter proportionella mot samtliga 40 personers fördelning på ålder.

### Uppgift 2:9

På två utbildningsprogram vid en högskola har man genomfört kursvärdering enligt ett fastställt formulär. Bland annat ställer man följande fråga: "Är du i huvudsak nöjd med din utbildning på programmet?" På ekonomprogrammet finns 60 studerande varav 42 personer besvarar frågorna och 24 svarar ja. På lärarprogrammet finns 50 studerande. Antalet svarande är 45 personer och 30 av dem svarar ja.

- Beräkna procentandelen nöjda (ja-svarare) bland de svarande på vart och ett av programmen samt hur många procent högre (eller lägre) som andelen nöjda är på lärarprogrammet jämfört med ekonomprogrammet.
- De som arbetar med ekonomprogrammet hävdar att jämförelsen är orättvis och att man mycket väl skulle kunna haft minst lika hög andel nöjda som "lärarna" om bara antalet svarande varit större. Utred om det ligger något i detta påstående.

### Uppgift 2:10

Man vill undersöka studieresultaten i ett visst ämne under hösten 2005. Bland annat är man intresserad av att se om studieresultaten skiljer sig mellan dem som antagits på gymnasiebetyg (betygsgruppen) och dem som antagits på resultat från högskoleprovet (provgruppen). Man undersöker hur många poäng som genomsnittligt producerats på kurser på olika nivå av studerande i de två grupperna. Resultatet av undersökningen redovisas i tabellen nedan. I tabellen anges alltså dels genomsnittligt antal poäng per studerande på olika kursnivå och dels (inom parantes) antalet studerande i var och en av grupperna. Vi ser alltså exempelvis att 22 personer i provgruppen som läst på lägre nivå i genomsnitt producerat 14.0 poäng.

Kursnivå	Betygsgrupp	Provgrupp
Lägre (A-B-nivå)	14.8 (48)	14.0 (22)
Högre (C-D-nivå)	17.9 (16)	17.6 (24)

- Beräkna med utgångspunkt från tabellen genomsnittligt antal poäng för studerande i betygsgruppen och motsvarande mått för de studerande i provgruppen.
- Jämför de båda grupperna (betygs- och provgrupp) med avseende på producerade poäng med hjälp av dels resultatet från a) och dels det som man ser i själva tabellen. Diskutera.

- c) Bilda standardvägda medeltal för producerade poäng, dels för studerande i betygsgruppen och dels för dem som tillhör provgruppen. Använd samtligas fördelning på kursnivå som underlag för standardvikterna.

### Uppgift 2:11

Vi betraktar en population om åtta hushåll vilka uppvisar en kostnad för elkonsumtion (i tusentals kronor) under en viss tidsperiod enligt följande: 9, 11, 27, 21, 17, 17, 22, 12.

Beräkna

- medelvärde, median och typvärde.
- undre och övre kvartil samt kvartilavstånd.
- beskriv materialet i ett lådagram (boxplot).
- varians och standardavvikelse.

### Uppgift 2:12

Man vill bilda sig en uppfattning om den genomsnittliga årliga körsträckan bland privatbilarna i ett land. Det handlar alltså om en mycket stor population. Man gör inledningsvis en provundersökning med ett slumpmässigt urval om  $n = 30$  privatbilägare och erhåller följande observationer på variabeln körsträcka i 1000-tal mil under perioden 1.7 2009 – 30.6 2010:

0.8 1.4 2.7 1.8 0.9 1.7 1.4 3.0 2.4 1.7 2.1 2.9 3.2 1.2 0.8 2.5 2.8  
1.8 1.9 2.6 3.2 3.1 0.4 1.1 1.4 1.9 1.8 2.8 0.9 2.9

- Beskriv de 30 körsträckorna i ett stam-och-blad-diagram.
- Beräkna den 20:e respektive 80:e percentilen i materialet om 30 körsträckor.
- De i uppräkningslistan ovan 12 första bilarna har kvinnliga ägare medan de 18 följande har manlig ägare. Beskriv körsträckorna för bilar med manliga respektive kvinnliga ägare i varsitt stam-och-blad-diagram.
- Beräkna mediansträcka respektive medelsträcka för såväl män som kvinnor.

## Uppgift 2:13

Man studerar lönerna på två vårdcentraler i en kommun. De 12 respektive 16 personer som är verksamma på de båda arbetsplatserna har i september 2009 följande löner i tusental kronor:

Vårdcentralen i C-mora

Pers nr: 1 2 3 4 5 6 7 8 9 10 11 12

Lön i tkr: 40 27 29 16 18 13 15 16 14 12 12 8

Vårdcentralen i D-lunda

Pers nr: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Lön i tkr: 23 19 17 27 27 23 13 16 13 12 9 33 18 6 6 7

- Beskriv lönefördelningen för de anställda vid vårdcentralen i C-mora med hjälp av dels en lämplig tabell och dels ett diagram lämpade för en föredragning för kommunledningen. Kommentera resultaten.
- Genomför motsvarande beskrivning med avseende på vårdcentralen i D-lunda.
- Beräkna lönesumman för vardera vårdcentralen samt medellönen per anställd på vardera arbetsplatsen. Kommentera resultaten.

För att kunna göra en rättvisande jämförelse mellan lönerna på de båda vårdcentralerna behöver man ta hänsyn till att inte alla anställda arbetar heltid. Följande gäller:

Heltid (40 tim/vecka)	3/4-tid (30 t/v)	1/2-tid (20t/v)
-----------------------	------------------	-----------------

C-mora: person nr 1-9	person nr 10-11	person nr 12
-----------------------	-----------------	--------------

D-lunda: person nr 1-9	person nr 10-13	person nr 14-16
------------------------	-----------------	-----------------

Lönerna som ges i början av uppgiften är alltså givna i ordning från heltidsanställda via 3/4- till halvtidsanställda. När man nu vill jämföra löneläget på de båda arbetsplatserna är det naturligt att göra jämförelsen på heltidstjänstbasis. Varje lön ska alltså omräknas till lön motsvarande heltidsanställning.

- Beräkna medellönen per anställd, om alla varit heltidsanställda, (halvtidslön multipliceras med 2 etc) för vardera vårdcentralen.

På vardera vårdcentralen finns anställda i tre olika kategorier, nämligen läkare, övrig vårdpersonal respektive övrig personal (lokalvårdare, läkarsekreterare m.fl.). De anställda fördelar sig på tjänst enligt nedan, där "personnummer" anges för respektive kategori.

	Läkare	Övrig vårdpersonal	Övrig personal
<b>C-mora:</b>	1, 2, 3	4, 5, 8, 9, 10, 11	6, 7, 12
<b>D-lunda:</b>	1, 4, 5, 6, 12, 13	2, 3, 7, 8, 10, 11, 14	9, 15, 16

e) Ta fram en tvåvägsindelad tabell med två kolumner svarande mot de två vårdcentralerna och tre rader svarande mot de tre tjänstekategorierna. I tabellcellerna ska såväl medellönen (för heltidstjänst) för de anställda som antalet anställda i var och en av de 6 cellerna finnas. Beräkna därefter standardvägda medellöner för var och en av vårdcentralerna. Välj standardvikter proportionella mot totala antalet i respektive tjänstekategori.

## Uppgift 2:14

I en undersökning vill man bl a studera storleken på bostäderna i ett bostadsområde. Som ett led i undersökningen sammanställer man följande tabell:

Antal bostadsrum (exkl kök och badrum)	Antal bostäder med vidstående antal rum	Procent
2	130	26
3	140	28
4	120	24
5	40	8
6	50	10
7	20	4
Totalt	500	100

- Beräkna medelantalet bostadsrum per bostad.
- Beräkna standardavvikelsen för antal bostadsrum.
- Beräkna medianantalet bostadsrum.
- Beräkna kvartilavstånd.

### Uppgift 2:15

En fristående kurs på totalt 20 poäng är uppdelad i 4 delkurser om vardera 5 poäng. De 30 personer, som varit registrerade på kursen har vid kursens slut uppnått varierande antal poäng. Fördelningen ser ut som följer:

Uppnått antal poäng, $X$	Antal personer
0	3
5	2
10	5
15	8
20	12

- Beräkna medelpoäng i gruppen om 30 personer.
- Beräkna standardavvikelsen i gruppen om 30 personer.
- Beskriv resultatet för de 30 personerna i ett lådagram.
- Beskriv materialet i ett stapeldiagram.

### Uppgift 2:16

Du vill jämföra lönenivån mellan två mindre företag. Eftersom lönen oftast hänger ihop med den anställdes utbildningsnivå, skaffar du information om dels månadslön och dels utbildning för var och en av de anställda. Du grupperar utbildningen i följande fyra kategorier, där varje anställd finns i en och endast en kategori

- F: Forskarutbildning
- H: Högskoleutbildning
- G: Gymnasieskola
- Ö: Övrig utbildning

Så här ser resultatet ut för de 15 anställda i A och de 12 anställda i B, när varje persons månadslön är angiven i antal tusen kronor:

Utbildning	Företag A	Företag B
F	32, 30, 25	31, 29, 27, 25
H	27, 23, 19, 19,	24, 20, 18, 18
G	18, 17, 16, 15, 14	16, 14
Ö	16, 15, 14	14, 14

- Beräkna medellönen för de 15 anställda i företag A och motsvarande mått för de 12 anställda i företag B.
- Bilda en tabell som lämpar sig som underlag för beräkning av standardvägda medeltal. Tabellen ska innehålla en kolumn för varje företag och en rad för varje utbildningsnivå. I varje cell ska du placera medellönen för de personer som finns i "cellen" samt hur många personerna är. Bilda därefter standardvägda medeltal för vardera företaget. Standardisera med avseende på utbildningskategori.
- Jämför resultaten under a) och b) och kommentera eventuella skillnader. Förklara varför det blir så i just det här fallet.

### Uppgift 2:17

I en kommun avser man byta ekonomiskt redovisningssystem och det kommer därför bli aktuellt att utbilda personalen på det nya systemet. Man tar fram två utbildningsmaterial avsedda i huvudsak för självstudier och testar de båda materialen på två av kommunens förvaltningar, det ena materialet på en förvaltning och det andra på den andra. Syftet är att komma underfund med vilken metod som ger kortast inlärningstid och därmed är minst resurskrävande. De anställda på de två förvaltningarna för noggrann bok över den tid som går åt för att ta sig igenom studiematerialet, så att man klarar av "slutprovet". Resultaten redovisas i tabellen, där man delat in de anställda på de båda förvaltningarna efter ålder i två grupper – Äldre och Yngre – och efter datorvana i två grupper med stor respektive liten vana. Följande resultat erhålls:

Genomsnittlig inlärningstid i timmar samt antal anställda (inom parentes)

Grupp	Material A, Förvaltning I	Material B, Förvaltning II
Äldre, stor vana	18 (22)	16 (16)
Äldre, liten vana	27 (18)	26 (32)
Yngre, stor vana	17 (36)	16 (22)
Yngre, liten vana	26 (14)	23 (40)

- a) Beräkna den genomsnittliga inlärningstiden för vart och ett av de båda utbildningsmaterialen.
- b) Beräkna standardvägda medeltal för inlärningstiden för vart och ett av de båda utbildningsmaterialen. Välj standardvikter proportionella mot samtligas fördelning på åldersgrupp och datorvana.
- c) Jämför resultaten under a) och b) och kommentera. Förklara varför det blir så här i just detta fall.

# Kapitel 3

---

## Uppgift 3:1

Man kastar en röd sexsidig tärning och registrerar antalet ögon upp. Följande händelser definieras:

A: Minst 4 ögon upp

B: Udda antal ögon upp

- a) Beskriv försökets utfallsrum och de båda händelserna med formelspråk.
- b) Bestäm sannolikheterna  $\Pr(A)$  och  $\Pr(B)$ .
- c) Bestäm sannolikheterna för komplementhändelserna.
- d) Är A och B disjunkta händelser?
- e) Är A och B oberoende händelser?

## Uppgift 3:2

En klass består utav 28 studenter, 16 kvinnor och 12 män. Klassen läser en kurs där det ingår ett projektarbete där man arbetar gruppvis om fyra studenter som dras slumpmässigt från hela klassen.

- a) På hur många olika sätt kan den första gruppen se ut, i avseende på de individuella studenterna?
- b) På hur många sätt kan den första gruppen se ut, i avseende på könsfördelningen? Alltså hur många olika antal män/kvinnor kan gruppen bestå utav.
- c) Vad är sannolikheten att gruppen består utav två kvinnor och två män?
- d) Stina och Lars har arbetat tillsammans i en tidigare kurs och vill gärna fortsätta samarbetet. Hur många sammansättningar av den första gruppen innehåller både Stina och Lars?



### Uppgift 3:3

Risken för att få en viss sjukdom är 1 på 100 000. Ett test som är skapad för att upptäcka detta är 99 procent effektiv att identifiera sjukdomen, det vill säga den ger ett korrekt positivt resultat i 99 procent av fallen personen har sjukdomen. Tyvärr ger testet också falska positiva resultat i fem procent av fallen då personen inte är sjuk.

- a) Vad är sannolikheten att testet ska ge ett positivt resultat överlag?
- b) Vad är sannolikheten att en person som fått ett positivt **resultat** verkligen har sjukdomen?
- c) Vad är sannolikheten att en person som fått ett **negativt** resultat verkligen har sjukdomen?

### Uppgift 3:4

Vi kastar två stycken sexsidiga tärningar oändligt antal gånger och beräknar summan av ögonen som tärningarna får.

- a) Beräkna sannolikheten för att få summan två på ett slumpmässigt kast.
- b) Vi kan anta att kasten är oberoende av varandra, beräkna sannolikheten att vi får summa två på två kast i rad.

# Kapitel 4

---

## Uppgift 4:1

Här presenteras data rörande hushållsstorlek (antal personer i hushållet) för de 40 hushåll som bor i ett visst villaområde:

2	6	4	6	3	4	4	4	4	4	3	7	4	4	5	5
2	5	3	3	5	3	4	5	5	3	6	2	4	4	3	5
1	4	5	4	5	1	5	4								

- Beskriv materialet i en frekvenstabell med såväl absoluta som relativa frekvenser och absoluta och relativa kumulerade frekvenser.
- Beskriv materialet i ett stolpdiagram.
- Beskriv materialet i ett lådagram.
- Beräkna medelvärde, varians och standardavvikelse i materialet.
- Beräkna andelen observationer som är större än 5.
- Beräkna andelen observationer som är större än eller lika med 5.
- Approximera fördelningen med en normalfördelning med medelvärde och standardavvikelse som i det faktiska materialet. Beräkna utifrån normalfördelningen andelen av observationerna som har värden som är större än 5 respektive större än eller lika med 5. Jämför med resultaten från punkt 5 och 6.
- Approximera som i föregående uppgift men beräkna andelen av observationerna som har värden som är mindre än 4 respektive motsvarande andel som har värden mindre än eller lika med 4. Jämför med motsvarande andelar i det faktiska datamaterialet.
- Beräkna (med hjälp av normalfördelningsapproximation) den 50:e percentilen i materialet (använd alltså normalfördelningstabellen). Jämför med värdet på 50:e percentilen som man får om man räknar på de 40 observationerna.

Här presenteras nu elförbrukningen i 1000-tal kilowattimmar för de 40 hushållen:

9	27	16	22	14	24	24	22	19	21	14	31	19	18	22	15
11	16	13	17	22	14	17	15	24	15	32	12	26	21	22	31
10	25	22	16	12	10	19	11								

- j) Klassindela materialet i fem klasser och beskriv det klassindelade materialet i en tabell och i ett histogram. Välj själv lämpliga klasstorlekar.
- k) Beräkna medianen, första och tredje kvartilen.
- l) Beskriv datamaterialet i ett lådagram.

### Uppgift 4:2

Följande månadslöner (tusentals kronor) utbetalas till de fem anställda på ett litet företag:

18 20 20 22 25

- a) Beräkna medellön, varians och standardavvikelse i materialet av fem observationer på variabeln  $X =$  månadslön.

Till jul utgår en "extralön i form av 10 tkr" till var och en av de anställda.

- b) Bilda de fem observationerna på variabeln årslön,  $Y = 10 + 12X$ .
- c) Beräkna årsmedellönen, variansen och standardavvikelsen i materialet av fem  $Y$ -observationer.
- d) Vilka relationer mellan måtten i deluppgift a) och måtten i deluppgift c) gäller?
- e) Vi tänker oss nu en godtycklig variabel  $X$  med medelvärde och standardavvikelse lika med  $\mu$  respektive  $\sigma$ . Beräkna medelvärdet, variansen och standardavvikelsen för variabeln

$$Z = \frac{(X - \mu)}{\sigma}$$

*Ledning: Utnyttja reglerna för linjära variabeltransformationer.*

### Uppgift 4:3

Antag att vi vet att andelen ensamstående bland samtliga kvinnor i 20-25-årsåldern är 30%. Vi väljer slumpmässigt (OSU) 5 kvinnor i 20-25-årsåldern och definierar variabeln:

$X$  = antal ensamstående bland de fem utvalda

- Beräkna sannolikheten att minst 2 kvinnor är ensamstående.
- Antag nu att vi i stället ska välja 50 kvinnor med OSU. Beräkna sannolikheten att vi då skulle få minst 20 ensamstående.

### Uppgift 4:4

Man kastar en röd sexsidig tärning och registrerar antalet ögon upp. Följande händelser definieras:

A: Minst 4 ögon upp

B: Udda antal ögon upp

Låt  $X$  vara antalet ögon upp vid ett tärningskast, och beräkna väntevärde och standardavvikelse för slumpvariabeln  $X$ .

*Ledning: Är alla de möjliga resultaten av ett tärningskast tillsammans ett stickprov eller en population?*

Man kastar nu två tärningar, vår tidigare röda och en likadan men grön. Beteckna antalet ögon upp på röd tärning med  $X$  och antalet ögon upp på grön tärning med  $Y$ . Man bildar slumpvariabeln

$$W = X + Y$$

och definierar händelsen

C: minst 10 ögon upp.

- Beskriv försökets utfallsrum och händelsen med formelspråk och bestäm sannolikheten för händelsen C.
- Beräkna väntevärde och standardavvikelse för slumpvariabeln  $W$ .  
*Ledning: Population eller stickprov?*

### Uppgift 4:5

Man genomför i tre olika situationer ett slumpmässigt försök som blir antingen lyckat (med sannolikheten  $p$ ) eller misslyckat (med sannolikheten  $1-p$ ). Försöket upprepas 4 gånger och man definierar slumpvariabeln:

$X$  = antalet lyckade försök

Här följer de tre situationerna

- Ett symmetriskt mynt kastas. Lyckat försök är krona medan klave är misslyckat.
- En vanlig symmetrisk sexsidig tärning kastas. Högst två ögon upp är lyckat försök medan tre eller fler ögon upp är misslyckat.
- Ur en påse med 50 kulor varav 30 röda och 20 blå väljs slumpmässigt en kula. Röd kula är lyckat försök medan blå är misslyckat.

Genomför för var och en av de tre situationerna följande:

- a) Bestäm sannolikhetsfördelningen för  $X$  och åskådliggör den i ett stolpdiagram.
- b) Beräkna sannolikheten för att  $X$  antar värdet 3 eller högre.
- c) Beräkna väntevärde och standardavvikelse för  $X$ .

Antag nu att man väljer 12 kulor ur en stor låda med 5000 kulor varav 3000 röda och 2000 blå.  $X$  betecknar antalet röda kulor bland de 12.

- d) Bestäm sannolikhetsfördelningen för  $X$  och åskådliggör den i ett stolpdiagram.
- e) Beräkna sannolikheten för att högst 3 av de 12 kulorna är röda.
- f) Beräkna väntevärde och standardavvikelse för  $X$ .
- g) Approximera fördelningen för  $X$  med en lämplig normalfördelning. Bestäm approximativt sannolikheten för att högst 3 av de 12 kulorna är röda.

# Kapitel 5

---

## Uppgift 5:1

Givet är en population med sex element med följande värden på variabeln  $X$ :

2      4      4      6      10      16

- a) Beräkna medelvärde och standardavvikelse i populationen.

*Ledning: Vi känner hela populationen.*

Antag nu att man gör ett slumpmässigt urval om  $n = 2$  element ur populationen och beräknar urvalsmedelvärdet  $\bar{x}$ . Urvalet görs med återläggning mellan dragningarna.

- b) Ta fram samplingfördelningen för  $\bar{x}$  genom att redovisa alla möjliga urval om  $n = 2$  enligt beskrivningen ovan.

*Ledning: Det totala antalet möjliga urval är 36.*

- c) Åskådliggör samplingfördelningen för  $\bar{x}$  i ett stolpdiagram.

- d) Beräkna medelvärde och standardavvikelse i samplingfördelningen för  $\bar{x}$ .

*Ledning: Räkna ut medelvärde  $\mu_{\bar{x}}$  och standardavvikelse  $\sigma_{\bar{x}}$  bland de  $\bar{x}$ -värden som tagits fram i punkt 2. Detta är samtliga möjliga urval om  $n = 2$  ur populationen och därmed också en population.*

- e) Visa att de relationer mellan populationsparametrarna och motsvarande parametrar i samplingfördelningen för  $\bar{x}$  som enligt litteratur bör gälla faktiskt också gäller.

*Ledning:  $\mu_{\bar{x}} = \mu$  och  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$*

- f) Är  $\bar{x}$  en väntevärdesriktig skattning av populationsmedelvärdet?

## Uppgift 5:2

Vår population utgörs av samtliga småföretag i ett land, och vi studerar kostnaderna för sjukfrånvaro, variabeln  $X$ , under en viss period. Vi tror att den genomsnittliga kostnaden är 60 tkr per företag och att standardavvikelsen är 16 tkr och arbetar därför med följande antagande: det gäller faktiskt att  $\mu = 60$  och  $\sigma = 16$ .

- a) Vi drar ett OSU om  $n = 400$  företag ur den mycket stora populationen och beräknar stickprovsmedelvärdet  $\bar{x}$ . Vad kan vi enligt centrala gränsvärdesatsen säga om fördelningen (samplingfördelningen) för  $\bar{x}$ ?

- b) Vad blir väntevärde och standardavvikelse i samplingfördelningen för  $\bar{x}$ ?

- c) Hur stor är sannolikheten att vårt observerade värde på  $\bar{x}$  hamnar mellan 58.43 och 61.57?
- d) Om det skulle visa sig att vårt antagande om värdet på  $\mu$  är felaktigt, hur skulle då resultaten under punkterna 1–3 påverkas?
- e) Om det skulle visa sig att vårt antagande om värdet på  $\sigma$  är felaktigt, hur skulle då resultaten under punkterna 1–3 påverkas?

### Uppgift 5:3

Ett livsmedelsföretag saluför nötter i 100-gramspåsar. Nöttsåarna levereras till grossist i lådor om 1200 påsar.

En grossist mottar en låda nöttsåar. Låt påsvikten betecknas med  $X$ . Vi antar att de 1200 påsvikterna i lådan kan betraktas som  $N(100, 1.5)$ .

- a) Hur stor är sannolikheten för att en ur lådan slumpmässigt vald nöttså väger mindre än 99 gram?
- b) Om grossisten slumpmässigt väljer två nöttsåar ur lådan, hur stor är då sannolikheten för att båda väger mindre än 99 gram?

*Ledning: Multiplikationssatsen för oberoende händelser.*

Vi återgår till företaget som saluför nötter. Man har där ett behov av att kontrollera att produktionsprocessen fungerar bra och genomför följande kontrollåtgärd: ur varje låda som ska säljas väljer man slumpmässigt 10 påsar bland de 1200 och beräknar medelvikten,  $\bar{x}$ . Om  $\bar{x}$  avviker från  $\mu$  (100 gram) med mer än 1 gram så kasseras hela lådan.

- c) Beräkna sannolikheten för att en låda kasseras givet att det faktiskt gäller att  $\mu = 100$  g och  $\sigma = 1.5$  g.

Företaget vill inte förkasta mer än högst 2% av de korrekta lådorna. Vi har sett i föregående uppgift att om man kräver att avvikelsen ska vara högst 1 gram så kommer en högre andel att förkastas.

- d) Beräkna storleken i gram på den avvikelse som man måste acceptera om högst 2% av de korrekta lådorna ska förkastas.
- e) Samtidigt som man vill ha en liten risk att förkasta en korrekt låda så vill man att chansen ska vara hög att förkasta en låda som inte är korrekt. Antag att i en viss låda är  $\mu = 103$  g och  $\sigma = 1.5$  g. Hur stor är chansen att man ska kassera lådan om det tidigare kravet om en avvikelse på högst 1 gram gäller?

*Ledning: Man kasserar alltså lådan om medelvikten av de 10 utvalda påsarna är mindre än 99 gram eller mer än 101 gram.*

- f) Hur stor är risken att lådan i punkten ovan inte kommer att förkastas?

### Uppgift 5:4

Vi återvänder till den population med sex element som användes i uppgift 5:1. Vi delar nu in populationen i två strata med de fyra minsta värdena, nämligen 2 4 4 6, i ett stratum och de två största, 10 16, i ett. Antag sedan att man väljer ett element ur vart och ett av de båda strata och bildar medeltalet av de båda observationerna.

- a) Ta fram samplingfördelningen för medelvärdet av de två observationerna genom att bilda alla möjliga kombinationer av ett element ur stratum 1 och ett element ur stratum 2. Åskådliggör samplingfördelningen i ett stolpdiagram.  
*Ledning: det totala antalet möjliga urval är 8.*
- b) Beräkna medelvärde och standardavvikelse i samplingfördelningen.
- c) Jämför samplingfördelningen med den som togs fram i uppgift 5:1. Kommentera.

### Uppgift 5:5

Vi tänker oss en totalundersökning riktad till de studerande och anställda vid en liten högskola. Man vill undersöka om det kan finnas underlag för ett bankkontor på campus. En mycket enkel enkät med några få frågor distribueras via postfack och internpost till samtliga studerande och anställda. 540 besvarade formulär utgör resultatet. Bland dessa har 405 svarat ja på frågan "anser du att det behövs ett bankkontor på campus?".

- a) Beräkna med utgångspunkt från undersökningsresultatet procentandelen i populationen som anser att det behövs ett bankkontor på campus.

Antag nu att vi får veta att det finns sammanlagt 980 anställda och studerande vid högskolan i fråga. Det finns alltså ett stort bortfall i undersökningen.

- b) Beräkna hur stor andelen ja-svar skulle ha kunnat bli som minst respektive mest, om hänsyn tas till bortfallet.
- c) Fundera lite kring vad man kan tro om bortfallet i detta fall. Är det att bedöma som slumpmässigt?



### Uppgift 5:6

I en fabrik produceras inlagda rödbetor på burk som säljs under beteckningen Bettans rödbetor, 1000 gram. Man anser sig veta att standardavvikelsen bland tillverkade burkar är 8 gram.

- Ungefär hur stor del av burkarna väger minst 1010 gram?
- Om man slumpmässigt väljer ut en burk, hur stor är då sannolikheten att den ska visa sig väga mindre än 995 gram?

Från en veckas produktion (ca 10 000 burkar) väljer man regelmässigt ett OSU om 125 burkar och väger dem. Om medelvikten bland de utvalda burkarna då avviker från 1000 gram med mer än 3 gram avbryter man produktionen och letar efter fel i produktionsprocessen. Vi betraktar produktionen en godtycklig vecka som vår undersökningspopulation.

- Om det är sant att populationen av burkar i genomsnitt väger just 1000 g, hur stor är då sannolikheten att få en medelvikt i urvalet som med minst 3 g avviker från 1000 g?
- Om i själva verket populationen av burkar i genomsnitt väger bara 998 g med en standardavvikelse som är 16 gram, hur stor är då sannolikheten att få en medelvikt i urvalet som med minst 3 g avviker från 1000 g?

### Uppgift 5:7

Givet är en population med sex element med följande värden på variabeln  $X$ :

2      4      4      6      10      16

Man gör urval om  $n = 2$  element med lika sannolikheter men utan återläggning och bildar urvalsmedeltalet  $\bar{x}$ .

- Ta fram samplingfördelningen för urvalsmedeltalet.
- Avbilda samplingfördelningen i ett stolpdiagram.
- Beräkna medelvärde och standardavvikelse i samplingfördelningen.
- Jämför de framräknade parametrarna med motsvarande värden i populationen.
- Jämför de framräknade parametrarna med motsvarande mått som togs fram i uppgift 5:1.
- Jämför samplingfördelningen enligt stolpdiagrammet (punkt b ovan) med motsvarande diagram i uppgift 5:1. Vilka slutsatser kan dras?

### Uppgift 5:8

Ett bageri bakar och säljer skorpor i helkilospåsar. En viss variation blir det dock i påsvikten. Räkna med att påsarna i genomsnitt väger 1000 gram med en standardavvikelse om 5 gram. Fördelningen för påsvikten anses vara ungefär normal.

- a) Ungefär hur stor del av skorppåsarna väger minst 1008 gram?
- b) Om man slumpmässigt väljer ut en påse, hur stor är då risken (sannolikheten) att den ska visa sig väga mindre än 995 gram?
- c) Hur mycket ska en skorppåse minst väga för att den ska tillhöra de 25% tyngsta paketen (övre kvartilen)?
- d) Antag nu att man slumpmässigt väljer ut 8 påsar ur en månads produktion. Beräkna sannolikheten för att medelvikten av de 8 påsarna överstiger 1002 gram.

# Kapitel 6

---

## Uppgift 6:1

Man vill undersöka (den vuxna) befolkningens inställning till det svenska medlemskapet i EU. Bland 120 tillfrågade slumpmässigt utvalda personer uppger 48 att de är negativa till det svenska medlemskapet.

- Bilda ett 95% konfidensintervall för andelen i hela populationen som är negativa till det svenska medlemskapet i EU.
- Ställ hypotesen att 45% av populationen är negativa till det svenska medlemskapet i EU mot den alternativa hypotesen att denna andel är mindre än 45%. Använd 5% signifikansnivå. Undersök om förutsättningarna för metoden är uppfyllda och beräkna testets  $p$ -värde.
- Om man i stället valt 1200 personer och 480 sagt sig vara negativa till det svenska medlemskapet, hur skulle då testet under punkt 2 påverkas? Undersök om förutsättningarna för metoden är uppfyllda och beräkna testets  $p$ -värde.

## Uppgift 6:2

Vi har en population bestående av elementen 5, 10, 20 och 65. Slumpmässigt och utan återläggning dras tre element. Man skattar populationsmedelvärdet  $\mu$  med ett intervall som sträcker sig från stickprovets minsta observation upp till den största. Bestäm intervallets konfidensgrad.

## Uppgift 6:3

Tänk dig att du får i uppgift att undersöka om den genomsnittliga viktökning som man kunnat se hos unga män under senare år har sin motsvarighet bland medelålders män. Låt oss anta att vi vet att bland medelålders män för 10 år sedan var 13.9% överviktiga. Antag så att du gör ett obundet slumpmässigt urval om 800 medelålders män, väger dem och finner att 15.6% är överviktiga enligt samma definition av övervikt som tidigare använts. På grundval av undersökningen vill du ta ställning till om andelen överviktiga bland medelålders män ökat eller ej under 10-årsperioden och genomför därför en statistisk hypotesprövning.

- Formulera nollhypotes och mothypotes.
- Förklara i relation till hypoteserna vad som menas med fel av första slaget (typ I) och fel av andra slaget (typ II).

- c) Antag att du genomför testet på 5 % signifikansnivå. Vad är innebörden i detta?
- d) Vad kan man säga om risken för fel av andra slaget i detta test?
- e) Vad menas med testets  $p$ -värde?
- f) Antag nu att testets  $p$ -värde blir 0.15 (15%). Vilken slutsats drar du av testet? Är resultatet signifikant eller ej?
- g) Har medelålders män ökat i vikt genomsnittligt sett eller ej? Förklara för uppdragsgivaren.

### Uppgift 6:4

Man vill genomföra en urvalsundersökning (OSU) i syfte att kartlägga TV-tittande hos svenska ungdomar. Särskilt vill man med ett 95% konfidensintervall skatta hur många timmar i genomsnitt som 12-åringar ser på TV en "vanlig" vecka. Man vill göra en intervallskattning av den genomsnittliga tiden,  $\mu$ , med en felmarginal om högst 0.5 timme.

- a) Antag att vi har anledning att räkna med att standardavvikelsen i populationen är cirka 8 timmar per vecka. Beräkna hur stort urval som man minst bör ta.
- b) Antag nu att undersökningen gjorts utan att man först gjort utredningen under a) och att man i ett OSU om 500 12-åringar erhöll medeltiden 28.6 timmar och standardavvikelsen 9.5 timmar. Bilda ett 95% konfidensintervall för  $\mu$ .

### Uppgift 6:5

Man genomför en totalundersökning av en population som innehåller  $N=1000$  personer i syfte att bestämma  $p$  – andelen "positiva" (ettor) – i populationen. Man får svar från  $N_S$  personer. Övriga  $N_b = (1000 - N_S)$  utgör bortfall. Beräkna för var och en av uppgifterna a)-d):

- dels bortfallsandelen,  $N_b/N$ ,
- dels ett intervall för  $p$ , sådant att  $p$  säkert ligger inom intervallet.  
*Ledning: Antag att bortfallet består av enbart positiva respektive enbart negativa för att täcka upp alla möjliga utfall av  $p$ .*

- a)  $N_S=800$  och antalet positiva bland svararna är 200
- b)  $N_S=800$  och antalet positiva bland svararna är 500
- c)  $N_S=400$  och antalet positiva bland svararna är 100
- d)  $N_S=400$  och antalet positiva bland svararna är 200

e) Sammanfatta vad bortfallet tycks betyda för slutsatserna i ett fall som detta.

### Uppgift 6:6

Vi har en population bestående av samtliga småföretag (här definierat som företag med färre än 10 anställda,  $N = \text{ca } 560\,000$ ) där vi studerar kostnaderna för sjukfrånvaro under en viss period. Vi drar ett OSU om  $n = 1400$  företag ur den mycket stora populationen och beräknar medelvärdet i urvalet till 61.5 tkr. Vi antar att vi vet att standardavvikelsen i hela populationen är 26 tkr.

1. Bestäm ett 95% konfidensintervall för  $\mu$ .
2. Pröva hypotesen att  $\mu = 60$  mot hypotesen att  $\mu$  är skild från 60. Välj 5% signifikansnivå.
3. Beräkna  $p$ -värdet för testet ovan.
4. Gör om testet men använd istället mothypotesen att  $\mu$  är större än 60.

Antag nu att urvalet bestod av  $n = 16$  företag i stället för  $n = 1400$  och att urvalsmedeltalet var 61.5 precis som ovan. Antagandet om standardavvikelsen i populationen gäller även här.

5. Gör om deluppgift 1-4 för  $n = 16$  och undersök hur svaren påverkas.

Låt oss nu tänka oss att vi släpper det något orealistiska antagandet att vi känner populationens standardavvikelse. I stället utnyttjar vi informationen från vårt urval om  $n = 1400$  företag ur populationen. Urvalet antas ha gett medeltalet 61.5 tkr och standardavvikelsen 24.8 tkr.

6. Beräkna ett 95% konfidensintervall för  $\mu$ .

# Kapitel 7

---

## Uppgift 7:1

Vi anknyter till uppgift 6:1, som handlade om inställningen till det svenska EU-medlemskapet. Bland 120 tillfrågade slumpmässigt utvalda personer uppgav 48 att de var negativa till det svenska medlemskapet.

En opinionsundersökning riktad till ett slumpmässigt urval (OSU) om 1080 personer för ett år sedan visade att andelen som då var negativa till det svenska medlemskapet var 35.5%.

1. Bilda ett 95% konfidensintervall för skillnaden i andelen "EU-negativa" i dag jämfört med för ett år sedan.
2. Undersök med hjälp av hypotesprövning om andelen "EU-negativa" bland den vuxna befolkningen i dag är högre än motsvarande andel för ett år sedan. Välj 5% signifikansnivå.

## Uppgift 7:2

300 studerande från tre olika utbildningsprogram har skrivit en och samma tentamensskrivning i engelska. Provet bedöms med vanlig tregradig skala, U, G och VG. De tre betygsnivåerna fördelar sig över de tre programmen enligt följande tabell:

Betyg	Program 1	Program 2	Program 3	Summa
Underkänd	40	40	20	100
Godkänd	53	35	12	100
Väl godkänd	42	36	22	100

Ange för vart och ett av följande tre påståenden om det är sant eller falskt med hjälp av statistiska test.

- a) Bland de underkända är andelen studerande som kommer från program 2 dubbelt så stor som andelen som kommer från program 3.
- b) Bland de godkända är andelen studerande som kommer från program 1 cirka 34 procentenheter högre än andelen som kommer från program 2.

*Ledning:  $d_0$  blir 0.34, och den poolade andelen kan inte användas. Variansen i teststatistikan blir därmed  $\sqrt{\frac{p_1*(1-p_1)}{n_1} + \frac{p_2*(1-p_2)}{n_2}}$*

- c) Bland de väl godkända är andelen studerande som kommer från program 2 14 procentenheter högre än andelen som kommer från program 3.

*Ledning: Samma som b).*

### Uppgift 7:3

Man vill bilda sig en uppfattning om hur mycket pengar som hushållen i Älvhaga lägger på apoteksvaror. Antag att man tänker sig en undersökning baserad på ett urval (OSU) om  $n$  hushåll ur hela populationen om 20 000 hushåll. Syftet är att ta reda på den genomsnittliga kostnaden per hushåll för apoteksvaror under den senaste månaden. Frågan är hur stort urval man ska välja.

I en liknande undersökning som genomförts i Östermo, en kommun av liknande storlek i en annan del av landet, fann man i sitt urval att den genomsnittliga kostnaden per hushåll för apoteksvaror under mars månad var 124 kr och standardavvikelsen var 80 kr.

- a) Bilda ett 95% konfidensintervall för den genomsnittliga kostnaden per hushåll i Östermo under mars, givet att urvalet bestod av 500 hushåll (OSU), där samtliga faktiskt svarade.
- b) Beräkna hur stort urval som man borde ta i Älvhaga om vi antar att standardavvikelsen i populationen om 20 tusen hushåll är just 80 kr och vi siktar på ett konfidensintervall med en felmarginal om högst 10 kr.

Antag nu att urvalet i Älvhaga faktiskt kom att bestå av 300 hushåll med ett medelvärde om 138 kr och en standardavvikelse om 88 kr. Undersökningen skedde i maj månad.

- d) Går det med hjälp av dessa data att fastställa om den genomsnittliga apoteksvarukostnaden är högre i Älvhaga än i Östermo? Genomför en formell hypotesprövning. Antag härvid att apoteksvarukonsumtionen är ungefär densamma i mars som i maj.

### Uppgift 7:4

I samband med en vinprovning avsmakas två sorters Riojavin av vardera tio vinkännare, som poängsätter vinerna på en skala från 1 (uselt) till 20 (utsökt).<sup>1</sup> Följande resultat erhålls:

Person	1	2	3	4	5	6	7	8	9	10
Vin A	12	11	14	12	13	10	15	16	14	14
Vin B	11	10	12	11	10	11	16	13	13	15

<sup>1</sup> Data hämtade från Vejde-Leander, Ordbok i statistik, sid 308.

Vi vill pröva hypotesen att vinerna är likvärdiga med två olika metoder. För vart och ett av fallen gäller att hypoteserna noggrant ska formuleras, testvariabel anges och slutsats dras.

- a) Vi antar att vi har två oberoende stickprov om 10 personer, ett bestående av de 10 personer som testat vin A och ett bestående av de tio personer som testat vin B. Går det att urskilja en skillnad mellan vinerna? Använd 5% signifikansnivå.
- b) Vi antar att tio personer dragits med OSU ur en stor population av vinkännare och att var och en av dem testat både vin A och vin B. Går det att urskilja en skillnad mellan vinerna? Använd 5% signifikansnivå.

### Uppgift 7:5

I en stor kommun vill man bland de anställda jämföra antalet sjukskrivningsdagar bland män och kvinnor. Man har data från två obundna slumpmässiga urval, dragna oberoende av varandra, ett bestående av män och ett av kvinnor bland de kommunanställda. För vardera urvalet har man för varje utvald person tagit reda på antalet frånvarodagar under den aktuella perioden. Man har därefter beräknat medelvärde respektive standardavvikelse bland observationerna. Följande resultat erhöles:

Grupp	antal observationer	medeltal	standardavvikelse
Kvinnor	18	36.9	18.6
Män	14	32.2	14.3

- a) Pröva hypotesen att den genomsnittliga sjukfrånvaron i hela kommunen är lika stor bland kvinnor som bland män. Använd dubbelsidig mothypotes.
- b) Går det att "fastställa" att kvinnorna i kommunen i genomsnitt troligen har större sjukfrånvaro än männen? Genomför en lämplig hypotesprövning. Glöm ej att formulera såväl hypoteser som slutsats.



### Uppgift 7:6

Man har gjort två urval av företag i en viss bransch, det ena bestående av 100 bland totalt 10000 småföretag och det andra bestående av 100 bland totalt 4000 större företag. Urvalen har gjorts helt oberoende av varandra. Syftet med undersökningen är att få ett grepp om företagens kostnader för sjukfrånvaro. Man har undersökt kostnaderna för november månad år 2005. Följande resultat erhöles:

	<b>Små företag</b>	<b>Större företag</b>
Medeltal, tkr	27.8	108.5
Standardavvikelse, tkr	15.1	38.2

- Beräkna ett 95% konfidensintervall för den genomsnittliga kostnaden för sjukfrånvaron bland större företag under november månad 2005. Tolka intervallet med ord.
- Beräkna ett 95% konfidensintervall för den genomsnittliga kostnaden för sjukfrånvaron bland samtliga 14 000 företag.
- Den situation som behandlas i denna övning är ett exempel på stratifierat urval. Försök bedöma om följande påstående är korrekt eller ej: "Vid stratifiering bör man försöka få så liten standardavvikelse som möjligt inom varje stratum".

# Kapitel 8

---

## Uppgift 8:1

Vi är intresserade av skolelevers TV-tittande (i timmar per vecka) och använder följande data som illustration. Vi har 12 flickor och 12 pojkar till vårt förfogande.

De tolv flickorna har följande värden: 14, 23, 30, 16, 18, 34, 8, 10, 19, 4, 14, 14

De tolv pojkarna har följande värden: 15, 25, 25, 18, 34, 40, 9, 10, 22, 6, 17, 19

- Beräkna ett 95% konfidensintervall för den genomsnittliga TV-tittartiden bland högstadieungdom och anta därvid att våra 24 ungdomar är slumpmässigt (OSU) valda bland samtliga (knappt 300 000) högstadieungdomar i landet. Formulera slutsatsen i ord.
- Pröva på 5% signifikansnivå hypotesen att det inte finns någon skillnad mellan könen när det gäller antal timmars TV-tittande. Anta härvid att vi har två slumpmässiga urval, som gjorts helt oberoende av varandra, ett om 12 flickor bland samtliga högstadieflickor och ett om 12 pojkar bland samtliga högstadiepojkar.
- Samma hypotes som i b) prövas men vi antar nu att det slumpmässigt valts 12 syskonpar om en bror och en syster. Syskonen är angivna i samma ordning. I det första syskonparet har flickan tittat 14 timmar och pojken 15 osv.

## Uppgift 8:2

Antag att vi har stratifierat en population av företag i två strata, ett innehållande 400 000 "enmansföretag" och ett innehållande 160 000 företag med mellan 1 och 9 anställda. Vi betecknar medelvärden i stratum 1 med  $\mu_1$  och medelvärdet i stratum 2 med  $\mu_2$ .

- Uttryck med formelspråk medelvärdet,  $\mu$ , i hela populationen om 560 000 småföretag.

Ur stratum 1 väljer man nu 1100 företag (OSU) och oberoende av detta väljer man 300 företag (OSU) ur stratum 2. För de sammanlagt 1400 utvalda företagen registreras kostnaden för sjukfrånvaro under den aktuella perioden. Antag att vi erhåller följande resultat:

$$n_1=1100, \bar{x}_1 = 30 \text{ tkr}, n_2= 300, \bar{x}_2 = 136 \text{ tkr}$$

- Bestäm en punktskattning av  $\mu$ .
- Hur stor andel av företagen har vi valt till urvalen ur stratum 1 respektive ur stratum 2?

Vi antar nu att vi vet att standardavvikelsen bland de 400 000 enmansföretagen är 10 tkr och att standardavvikelsen bland de 160 000 flermansföretagen är 30 tkr.

- d) Beräkna ett 95% konfidensintervall för  $\mu$ .
- e) Om vi i stället hade valt 700 företag ur vardera stratomet och råkat få just de medeltal som vi fick förut (30 respektive 136), vad skulle då bli annorlunda jämfört med konfidensintervallet ovan? Beräkna ett 95% konfidensintervall och utred skillnaden.

### Uppgift 8:3

Året innan (år  $t - 1$ ) gjordes en liknande undersökning som den som gjordes nu (år  $t$ ) i uppgift 8:1. Man undersökte då – det tidigare året – ett urval (OSU) om 1250 småföretag och fick ett urvalsmedelvärde om 58 tkr och en standardavvikelse i urvalet om 17.2 tkr.

- a) Bilda ett 95% konfidensintervall för den genomsnittliga sjukfrånvarokostnaden det tidigare året (år  $t-1$ ). Jämför detta konfidensintervall med det som gjordes i uppgift 8:2. Vad skiljer dem åt?
- b) Bilda ett 95% konfidensintervall för skillnaden i genomsnittlig sjukfrånvarokostnad mellan innevarande år (år  $t$ ) och året före (år  $t - 1$ ). Vi antar då att undersökningen från uppgift 3:B i urvalet om 1400 småföretag gav till resultat urvalsmedelvärdet 61.5 tkr och standardavvikelsen 24.8 tkr.
- c) Undersök om det kan anses fastställt att den genomsnittliga kostnaden i populationen är högre i år (år  $t$ ) än förra året (år  $t - 1$ ). Formulera hypoteser, bilda testvariabel, välj signifikansnivå och beräkna kritiskt värde.

Vi tänker oss nu en alternativuppläggning av undersökningen sådan att vi valde ut (OSU) ett antal företag förra året, tog reda på deras sjukfrånvarokostnader under en viss period och därefter undersökte samma företag ett år senare och tog reda på motsvarande kostnader i år. Vi tänker oss ett urval om bara 12 företag (för att underlätta räknandet) med följande resultat:

Företag	1	2	3	4	5	6	7	8	9	10	11	12
Kostnad år 1 (tkr)	28	27	58	88	65	88	39	69	82	32	29	43
Kostnad år 2 (tkr)	36	34	66	94	70	86	44	78	84	36	35	45

- d) Ta ställning till samma hypotes som under punkt c).

*Ledning: Tänk igenom vad som skiljer de båda undersökningsuppläggingarna åt.*

### Uppgift 8:4

Man har gjort två urval av företag i en viss bransch, det ena bestående av 100 bland totalt 10 000 småföretag och det andra bestående av 100 bland totalt 4000 större företag. Urvalen har gjorts helt oberoende av varandra. Syftet med undersökningen är att få ett grepp om företagens kostnader för sjukfrånvaro. Man har undersökt kostnaderna för september månad ett visst år. Följande resultat erhöles:

	Små företag	Större företag
Medeltal, tkr	27.8	108.5

Vi antar att standardavvikelsen för sjukkostnaderna är känd inom vardera stratumet. I stratum 1 är standardavvikelsen 15 tkr och i stratum 2 är den 75 tkr.

- Beräkna ett 95% konfidensintervall för den genomsnittliga kostnaden för sjukfrånvaron bland större företag under september det aktuella året. Uttala resultatet så att icke statistikkunniga förstår vad undersökningen visat.
- Beräkna ett 95% konfidensintervall för den genomsnittliga kostnaden för sjukfrånvaron bland samtliga 14000 företag under den aktuella månaden.
- Denna situation är ett exempel på stratifierat urval. Ta ställning till följande påstående och ange om det är sant eller ej med kort motivering: Vid stratifiering bör man försöka få så liten standardavvikelse som möjligt inom varje stratum (delpopulation).

### Uppgift 8:5

Inom en mils radie från ett nyöppnat köpcenter bor 4000 hushåll. Man överväger att öppna ett apotek i köpcentret och för att bedöma kundunderlaget görs ett urval av hushåll (OSU) bland de 4000 och dessa kontaktas. Man ställer frågan "Räknar du med att ditt hushåll kommer att köpa merparten av sina apoteksvaror vid det eventuellt nya apoteket?" Av 120 svar är 72 ja-svar.

- Beräkna ett 95% konfidensintervall för andelen ja-svarare bland de 4000 hushållen (alltså den andel ja-svar som man skulle fått om man kunnat göra en totalundersökning av populationen).

Inom en radie om fyra mil från det nya köpcentret bor 20 000 hushåll (alltså har 16 000 hushåll mellan en och fyra mil till köpcentret.). Man gör ett urval bland dessa 16 000 hushåll och ställer samma fråga som tidigare. Bland 400 svar är 28 ja-svar.

- Använd nu data från båda undersökningarna och skatta andelen ja-svar bland de 20 000 hushållen med ett 95% konfidensintervall.
- Bilda också ett 95% konfidensintervall för antalet ja-svarande hushåll i populationen.

### Uppgift 8:6

Man undrar hur stor del av Norrköpings vuxna befolkning som är positiva till vårt EU-medlemskap. Av 1000 slumpmässigt utvalda personer uppger 520 att de är EU-positiva.

- Bilda ett 95% konfidensintervall för andelen EU-positiva bland Norrköpings vuxna befolkning.
- Bilda ett 95% konfidensintervall för antalet EU-positiva bland Norrköpings vuxna befolkning, som antas bestå av cirka 100 tusen personer.

### Uppgift 8:7

Man genomför en postenkätundersökning på ett OSU om 2000 villaboende hushåll ur en mycket stor population i syfte att dels skatta den genomsnittliga boendekostnaden (nettokostnad inklusive el och vatten och efter avdrag för skattereduktion, bostadsbidrag etc.) i september ett visst år och dels andelen av hushållen som har direktverkande el för uppvärmning. Man fick svar från 1200 hushåll och alltså ett bortfall om 40%. Av de svarande uppgav 390 hushåll att de hade direktverkande el och för boendekostnaden bland svararna gällde ett medelvärde om 7200 kr och en standardavvikelse om 2100kr.

- Beräkna lägsta respektive högsta värde på punktskattningen för andelen hushåll med direktverkande el i populationen genom antaganden om bortfallet.

Antag att man gjorde ett OSU om 80 hushåll ur bortfallet och att man med hjälp av personlig intervju lyckades få svar från samtliga 80. Den genomsnittliga boendekostnaden bland de 80 befanns vara 9600 kr och standardavvikelsen blev 1800 kr. Antalet hushåll med direktverkande el bland de 80 befanns vara 60.

- Beräkna en punktskattning för den genomsnittliga boendekostnaden i september i populationen.
- Gör ett försök att beräkna en felmarginal – att användas i ett 95% konfidensintervall – till punktskattningen under b).

*Ledning: Betrakta populationen som stratifierad i ett svarsstratum och ett bortfallsstratum. Den relativa stratumstorleken för respektive stratum skattas med hjälp av urvalets fördelning på svarare och bortfall.*

### Uppgift 8:8

Vi tänker oss en stor population om sammanlagt 10 000 företag som vart och ett gjort vissa investeringar under en given period. För enkelhets skull tänker vi oss att det handlar om 8000 mindre företag och 2000 större. Vi avser med hjälp av ett stratifierat urval om sammanlagt 400 företag bilda ett 95% konfidensintervall för de genomsnittliga investeringarna per företag i populationen.

- Beräkna felmarginalens storlek vid "kändsigmavariant" om vi dras OSU om  $n = 200$  från vardera stratum och antar att standardavvikelseerna är 7.0 respektive 20.0.
- Hur ska observationerna fördelas på de två strata om proportionell allokering ska användas?

Antag att vi slumpmässigt väljer 320 mindre företag och 80 större och erhåller följande resultat:

Stratum 1     $n = 320$      $\bar{x} = 16.5$      $s = 6.9$

Stratum 2     $n = 80$      $\bar{x} = 415.8$      $s = 22.4$

- Beräkna felmarginalens storlek om vi vill skatta medelvärdet per företag i hela populationen.
- Beräkna ett 95% konfidensintervall för de totala investeringarna som gjorts i populationen under den aktuella perioden.
- Fundera kring hur om man skulle kunna fördela de totalt 400 observationerna på de båda strata på ett ur "felmarginalsynpunkt" mera förnuftigt sätt.

# Kapitel 9

---

## Uppgift 9:1

En vara säljs i tre olika typer av enkilosförpackningar, typerna A, B och C. Man undrar om tonåringars preferenser för olika förpackningar skiljer sig från vad som gäller för äldre personer.

Bland 200 slumpmässigt valda tonåringar är preferenserna: 88 personer föredrar A, 58 personer föredrar B, 54 personer föredrar C.

Bland 300 slumpmässigt valda äldre personer är preferenserna: 162 personer föredrar A, 72 personer föredrar B och 66 föredrar C.

- Bilda en flerfältstabell med absoluta frekvenser med en kolumn för yngre och en för äldre personer och tre rader, en för varje förpackningstyp. Tabellen ska alltså visa hur de 500 personerna fördelar sig på ålder och preferens.
- Bilda en motsvarande tabell med den procentuella fördelningen för vardera åldersgruppen (kolumnerna summerar till 100% var). Bedöm om det finns någon skillnad mellan åldersgrupperna när det gäller preferenser för förpackningstyp.
- Analysera frågeställningen med hjälp av chitvåtest. Formulera hypoteser, beräkna förväntade frekvenser och observerat värde på testvariabeln och dra slutsats med ord. Använd 5% signifikansnivå.

## Uppgift 9:2

Man ville undersöka om det fanns någon åldersskillnad när det gäller val av nyhetsprogram i TV. Femhundra personer hade uppgett vilket program de föredrog av Rapport, Aktuellt och Nyheterna i TV4. Var och en av de tillfrågade uppgav också sin ålder och man har därmed kunnat göra följande tabell, som beskriver hur de sammanlagt 500 personerna fördelade sig på ålder och program.

TV-program	Under 20 år	20-39 år	40 år och mer
Rapport	76	86	58
Aktuellt	48	66	56
Nyheter i TV4	26	48	36

- Beskriv marginalfördelningen för ålder i en enkel frekvenstabell med relativa frekvenser.
- Beskriv i en tabell lämpad för att jämföra de tre åldersgrupperna hur var och en av de tre åldersgrupperna procentuellt fördelar sig på TV-program. Vilken av grupperna är mest "Rapportpositiv"?
- Undersök om de observerade skillnaderna är tillräckligt stora för att ett chitvåtest ska ge signifikant resultat. Ange hypoteser, beräkna testvariabelns värde och dra slutsats. Använd 5% signifikansnivå.

### Uppgift 9:3

600 personer i åldrarna 21-60 år har svarat på frågor rörande TV-program. I tabellen redovisas hur de svarande fördelar sig på sådana som svarar ja respektive nej på frågan "Brukar du titta på de svenska uttagningarna till schlagerfestivalen?":

Kön	Ålder	Antal som brukar titta	Antal som ej brukar titta
Kvinnor	21-40	82	28
Kvinnor	41-60	80	60
Män	21-40	112	68
Män	41-60	70	100

- Hur ser sambandet mellan ålder och "tittarvanor" (tittar/tittar ej) ut? Ställ upp en lämplig tabell för att studera sambandet. Tolka sambandet i ord.
- Hur ser sambandet mellan kön och ålder ut i gruppen av personer? Ställ upp en lämplig tabell för att studera detta samband och tolka sambandet i ord.



### Uppgift 9:4

280 slumpmässigt valda personer ur en stor population delas in efter ålder och åsikt om trafiken i Linköpings stadskärna på sätt som framgår av följande tabell:

Åsikt	Yngre	Äldre
Trafiken mycket störande	48	72
Trafiken något störande	32	72
Trafiken ej störande	40	16
Summa	120	160

Hypotesen att det föreligger ett beroende mellan åldersgrupp och åsikt om trafiken ska prövas med chitvåtest på 5% risknivå.

- Beräkna det aktuella chitvåvärdet och ta ställning till om hypotesen kan förkastas eller ej.
- Beräkna ett 95% konfidensintervall för procentandelen personer i populationen som anser att trafiken ej är störande.

### Uppgift 9:5

Betrakta situationen i uppgift 9:4 som om vi hade två oberoende OSU (ett ur populationen äldre och ett ur populationen yngre).

Undersök om andelen som tycker att trafiken är mycket störande är signifikant högre bland äldre än bland yngre. Testet ska genomföras på 1% signifikansnivå. Formulera hypoteser, redovisa testvariabeln och beräkna  $p$ -värde. Formulera slutsatsen i ord.

# Kapitel 10

---

## Uppgift 10:1

Tio slumpmässigt utvalda högskolestuderande med praktikarbete sommaren 2010 beskrivs med avseende på  $X$ : antal studieår vid högskola och  $Y$ : månadslön under sommaren i tkr. Följande data föreligger:

Person	1	2	3	4	5	6	7	8	9	10
$x$	1	2	2	3	3	3	4	4	4	4
$y$	11	11	14	12	14	16	13	14	15	18

- Beskriv materialet i ett spridningsdiagram.
- Beräkna  $b_0$  och  $b_1$  i regressionsekvationen  $y = b_0 + b_1x$ .
- Rita in regressionslinjen i diagrammet från punkt 1.
- Beräkna korrelationskoefficienten mellan  $x$  och  $y$ .
- Bestäm förklaringsgraden.
- Ta fram residualerna och åskådliggör dem på lämpligt sätt.
- Sammanfatta med ord vad man kan säga om sambandet mellan studietid och månadslön.

## Uppgift 10:2

Vi använder här samma datamaterial som i uppgift 10:1 och som handlar om tio slumpmässigt utvalda högskolestuderande med praktikarbete under sommaren.  $X$  har definierats som antal studieår vid högskola och  $Y$  som månadslön under sommaren i tkr.

- Pröva hypotesen  $\beta = 0$  med ett dubbelsidigt test på 5% signifikansnivå.
- Bilda ett 95% konfidensintervall för den genomsnittliga månadslönen för personer med 3 års högskolestudier.
- Gör ett prognosintervall för en person med 3 års högskolestudier.
- Gör ett prognosintervall för en person med 5 års högskolestudier. Jämför intervallet med intervallet under föregående punkt.
- Antag att personerna 1, 4, 5 samt 7 och 8 är studerande vid medellånga vårdutbildningar medan övriga personer är studerande vid tekniska högskolor. Konstruera en tredje variabel,  $Z$ , som en indikatorvariabel som skiljer fakulteterna åt.
- Beskriv i ett vanligt spridningsdiagram sambandet mellan de tre variablerna genom att använda olika symboler för de med  $Z = 1$  respektive  $Z = 0$ . Vad tycks fakultet betyda?

- g) Beskriv i symboltermer den modell som skulle kunna användas för att analysera sambandet mellan både studietid och fakultet å den ena sidan och månadslön å den andra.

### Uppgift 10:3

Ett företag tillverkar och säljer bland annat mysli. På marknadsavdelningen vill man närmare undersöka konsumtionsvanorna när det gäller mysli. Man samlar in uppgifter från ett urval av kunder. Vi låtsas nu att urvalet bestod av bara 10 kunder, så att vi får ett litet datamaterial som vi kan räkna på. Från urvalet samlar man in uppgift om de 10 kundernas familjestorlek, liksom om man har barn under 18 år eller ej (den informationen används ej här) och myslikonsumtion i kg per månad. Följande data erhålls:

Kundnummer	Familjestorlek	Myslikonsumtion
1	1	1
2	5	1
3	6	5
4	2	1
5	1	2
6	3	1
7	5	2
8	2	3
9	3	3
10	2	1

- a) Beskriv samband mellan familjestorlek och myslikonsumtion i ett spridningsdiagram. Låt familjestorlek vara  $x$  och myslikonsumtion  $y$ .
- b) Beräkna korrelationskoefficienten mellan myslikonsumtion och familjestorlek.
- c) Vad kan man säga om styrkan hos sambandet? Beräkna förklaringsgraden och försök beskriva med ord vad man kommit fram till. Verkar sambandet vara starkt?

Man hittar sedan ett fel i datafilen. Man har förväxlat konsumtionssiffran för kund nr 2 och nr 8. Det skulle alltså vara så att kund nummer 2 har  $x = 5$  och  $y = 3$  medan kund nr 8 har  $x = 2$  och  $y = 1$ .

- d) Beräkna den rätta korrelationskoefficienten.
- e) Beskriv kort vad du lärt dig av att man rättade felet i data.
- f) Sammanfatta vad man lärt om sambandet mellan familjestorlek och myslikonsumtion.

### Uppgift 10:4

Man vill undersöka sambandet mellan nedlagd studietid och tentamensresultat på en nationalekonomitenta med 20 studerande, vilka samtliga uppgett nedlagd studietid på kursen i timmar per vecka. För var och en av de studerande har man noterat tentaresultatet i antal poäng på en 50-gradig skala. Man har därefter genomfört en regressionsanalys med tentaresultatet som beroende och arbetstiden som oberoende variabel och dessutom tagit fram beskrivande mått för de båda variablerna enligt nedan.

Betrakta de 20 studerande som vore de slumpmässigt valda bland samtliga studenter på den aktuella kursen under den senaste treårsperioden (cirka 500 studenter).

Descriptive Statistics						
Variable	N	Mean	Median	TrMean	StDev	SE Mean
Arbetstid	20	31.55	30.00	31.33	10.47	2.34
Resultat	20	30.75	29.00	30.78	12.23	2.74
Variable	Minimum	Maximum	Q1	Q3		
Arbetstid	15.00	52.00	22.75	39.50		
Resultat	11.00	50.00	22.25	41.00		

  

Regression analysis				
The regression equation is				
Resultat = - 0.95 + 1.00 Arbetstid				
Predictor	Coef	StDev	T	P
Constant	-0.948	4.662	-0.20	0.841
Arbetsti	1.0047	0.1406	7.15	0.000
S = 6.417		R-Sq = 73.9%		

- Tolka i ord lutningskoefficienten i ovanstående regressionsekvation.
- Bestäm ett 95% konfidensintervall för den genomsnittliga arbetstiden per person i populationen
- Beräkna korrelationskoefficienten mellan arbetstid och resultat.
- Pröva hypotesen att det inte finns något (linjärt) samband mellan arbetstid och tentaresultat. Välj 5% signifikansnivå.
- Kommentera antagandet om att se de 20 personerna som ett slumpmässigt urval.

### Uppgift 10:5

Vi arbetar här med data hämtade från en (gammal) undersökning av sambandet mellan inkomst, banktillgångar och nysparande (sparande under året) i hushåll. Sorten på de ekonomiska variablerna är tusental kronor. Materialet som använts här består av 9 hushåll och uppgifterna har analyserats i Minitab. Nysparande betraktas i regressionsanalyserna som beroende variabel.

I tabellen nedan finns samlad information från tre enkla regressionsanalyser med nysparande som beroende variabel och var och en av de tre övriga som förklaringsvariabel i varsin analys.

Variabel	Medelvärde	Standardavvikelse	Korrelation med nysparande	$\sum e^2$	$\sum (X - \bar{X})^2$
Nysparande	7.33	2.828			
Inkomst	80.0	15.81	+0.839	19.011	1999.64
Bankmedel	90.0	47.4	- 0.680	34.406	17974.08
Antal barn	2.556	1.333	- 0.685	33.972	14.22

Här följer en regressionsanalys med en förklaringsvariabel.

### Regression Analysis

The regression equation is

$$\text{Nysparande} = 11.0 - 0.0406 \text{ Bankmedel}$$

Predictor	Coef	SE Coef
Constant	10.983	1.660
Bankmedel	-0.04056	0.01652

S = 2.217      R-Sq = 46.3%

- Tolka med ord lutningskoefficienten i ovanstående regressionsekvation.
- Pröva hypotesen att det inte finns något linjärt samband mellan bankmedel och nysparande.

Det skulle ju vara möjligt att göra tre olika regressionsanalyser med nysparande som beroende variabel och inkomst eller bankmedel eller antal barn som förklarande variabel. (Du har all information du behöver för de följande deluppgifterna i tabellen.)

- Vilken av dessa tre analyser skulle du välja om du var tvungen att satsa på bara en? Motivera.
- Bilda ett 95% konfidensintervall för lutningskoefficienten för förklaringsvariabeln som du valt i deluppgift c. Tolka intervallet.

# Svar till vissa uppgifter

---

## 2:11

- a) 17, 17, 17
- b) 11.5 och 21.5, 10
- c) De fem talen är 9, 11.5, 17, 21.5, 27
- d) 38 respektive 6.16

## 2:12

- a) 1.15 respektive 28.5
- b) Kvinnliga: 1.9 respektive 1.75 tusen mil, manliga: 2.02 respektive 1.9 tusen mil

## 2:13

- a) 220 tkr respektive 269 tkr, 18.33 tkr respektive 16.81 tkr
- b) 19.67 respektive 19.50
- c)

Tjänstekategori	C-mora, medellön + (antal)	D-lunda, medellön + (antal)
Läkare	32 (3)	28 (6)
Övr vårdpersonal	16 (6)	15 (7)
Övrig personal	14.7 (3)	13 (3)

Standardvägda medeltal 20.86 respektive 18.75

## 2:14

- a) 3.6
- b) varians 2.0 och standardavvikelse 1.414
- c) 3

d) kvartilerna 2 och 4, kvartilavståndet 2.

### 2:15

a) 14

b) 6.51

c) 0, 10, 15, 20, 20 (de fem måtten)

### 2:16

a) 20 tkr respektive 20.8 tkr

b) 21.0 tkr respektive 19.7 tkr, med standardvikter proportionella mot samtliga 27 personers fördelning på utbildning.

### 2:17

a) 20.6 och 21.5

b) 22.1 och 20.4

c) I a) får material B längre genomsnittlig tid än material A (21.5 respektive 20.6) trots att material A har högre genomsnittstid på varje rad i tabellen. Detta beror på att material A testas på en majoritet av personer med stor vana (22+36=58 personer av 90) och som därför får kort inläringstid medan material B testas på en majoritet av ovana (72 av 110) som får lång inläringstid. När samma vikter används för båda materialen, som i b) slår det faktum att A ger längre genomsnittstider än B i varje grupp igenom och ger A ett större standardvägt medeltal än B (22.1 jmf med 20.4).

### 4:3

a) 0.47

b) 0.0848

### 5:8

a) 5.48%

b) 0.16 (15.87%)

c) 1003.35 g (motsvarar z-värdet 0.67)



d) 0.13 (12.92%)

### 6:4

a) 984

b) 27.77 till 29.43

### 6:5

a) 0.2, 0.2 till 0.4

b) 0.2, 0.5 till 0.7

c) 0.6, 0.1 till 0.7

d) 0.6, 0.2 till 0.8

e) Intervalllängden tycks vara lika med bortfallets storlek.

### 7:6

a) 101 till 116

b) 47.86 till 53.86

c) Påståendet är sant. I formeln för felmarginalen kan man se att ju mindre stratumvarianser desto kortare felmarginal.

### 8:8

a) 0.95(38)

b) 320 respektive 80

c) 1.15(31)

d) 952100 till 975100

e) Eftersom standardavvikelsen är så mycket större i stratum två skulle det förmodligen löna sig (ge kortare felmarginal) om man tog lite fler ur det stratomet än de 80, som vi räknat med i c) och d). Man kan ju pröva sig fram lite och se om inte drygt 200 små och knappt 200 stora kanske vore att föredra.

#### 9:4

a) Beräknat chitvåvärde=25.29, kritiskt värde=5.99. Nollhypotesen förkastas.

b)  $0.2 \pm 0.047$

#### 9:5

Testvariabelns beräknade värde 0.86, kritiskt värde 2.33. Nollhypotesen förkastas ej.

#### 10:5

a) För varje extra 1000-tal kronor i bankmedel så minskar nysparandet enligt modellen med 40.6 kronor.

b) Det observerade t-värdet är  $-2.455$  och tabellvärdet  $-2.365$ .  $H_0$  förkastas alltså.

c) Ett möjligt svar (i komprimerad form): Den med den bästa förklaringsgraden, nämligen modellen med inkomst. Denna modell känns också mest logisk.

d)  $0.15 \pm 0.088$ .