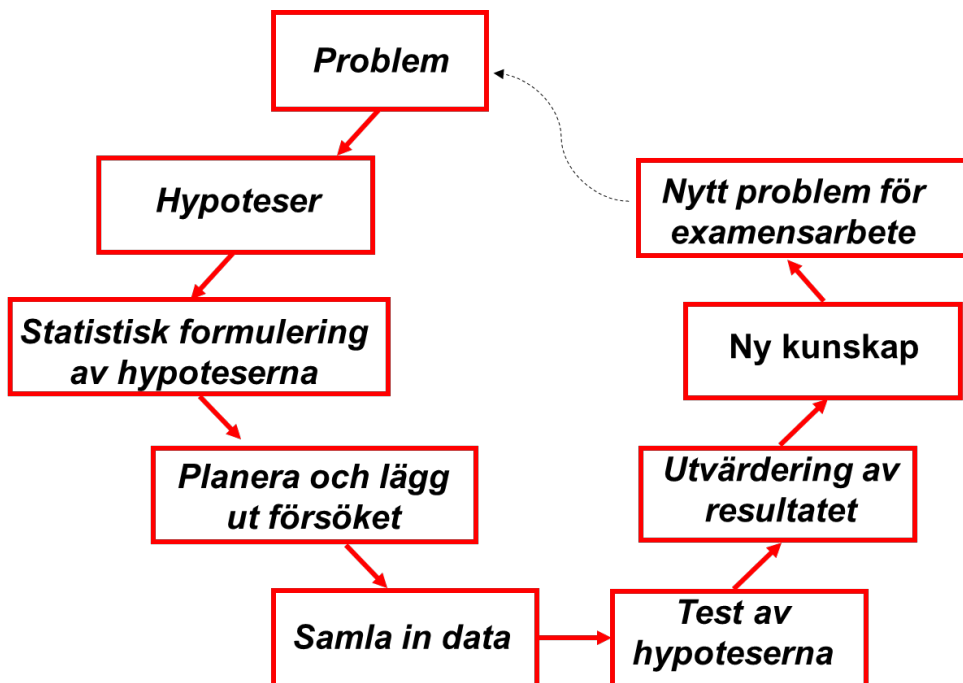


Dags för en kvantitativ analys!



Förord

Bakgrunden till det här projektet och kompendiet är att trädgårdsingenjörer har ganska lite statistik i sin utbildning och därför hamnar i svårigheter när de kommer till att göra ett examensarbete som fordrar en kvantitativ analys.

Genom de projekt vid LTV-fakulteten som kallas Best Practice fick vi tillfälle att göra ett material som skulle hjälpa till vid dessa kvantitativa analyser.

Detta kompendium kan läsas från första till sista sidan men det går också att gå in på enskilda avsnitt eller kapitel för att se mer i detalj hur du gör i vissa delar av examensarbetet.

Det finns också material på Canvas som kompletterar det här materialet och där kan man också hitta det Minitabkompendium som refereras till på några ställen.

Alnarp våren 2022

Jan-Eric Englund

Innehåll

1	Problemformulering.....	1
1.1	Kvantitativ analys.....	1
1.2	Arbetsgång för examensarbetet.....	1
1.3	Vetenskap och beprövad erfarenhet.....	2
1.4	Experiment eller observationsstudie.....	2
1.5	Statistiska hypoteser.....	4
1.6	Replikat.....	4
1.7	Population.....	6
2	Försöksplanering.....	9
2.1	Slumpmässigt stickprov.....	9
2.2	Randomisering.....	10
2.3	Frågeställningar.....	11
2.4	Två behandlingar.....	12
2.5	Mer än två behandlingar.....	13
3	Insamling av data.....	15
3.1	Generella råd.....	15
3.2	Olika typer av variabler.....	15
3.3	Observationer i datorn.....	17
3.4	Några vanliga frågor.....	18
4	Analys av data.....	19
4.1	Inledning.....	19
4.2	Medelvärde.....	19
4.3	Median.....	22
4.4	Standardavvikelse.....	23
4.5	Standard Error (medelfel).....	25
4.6	Varians.....	25
4.7	Stapeldiagram.....	26
4.8	Punktdiagram.....	27
4.9	Lådagram.....	28
4.10	Linjediagram.....	30
4.11	Histogram.....	30
4.12	Cirkeldiagram.....	31
4.13	Dataanalys i Excel.....	32
5	Vetenskapligt förankrade slutsatser.....	34
5.1	Definition av p-värde.....	34

5.2	Arbetsgång vid hypotestest	35
5.3	Ett stickprov.....	39
5.4	Stickprov i par.....	43
5.5	Två oberoende stickprov.....	47
5.6	Konfidensintervall för medelvärdet.....	50
5.7	Mer än två behandlingar	53
5.8	Variansanalys utan blockindelning.....	54
5.9	Variansanalys av ett blockförsök.....	57
5.10	Variansanalys med flera faktorer.....	60
5.11	Eftertester i variansanalys.....	63
5.12	Fishers metod	65
5.13	Tukeys metod	66
5.14	Ytterligare metoder	68

1 Problemformulering

1.1 Kvantitativ analys

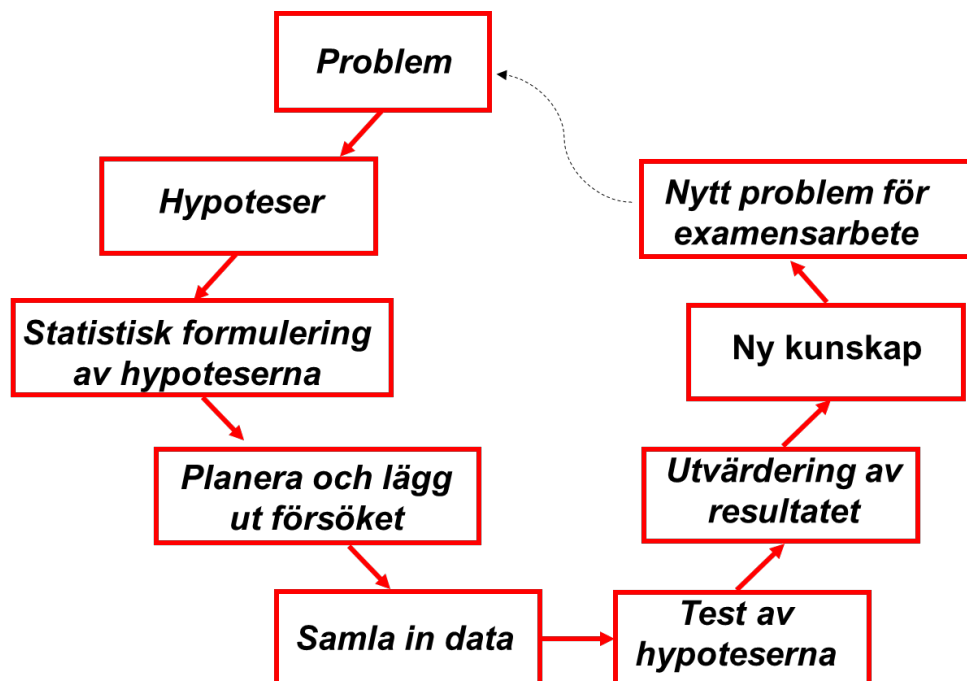
Detta kompendium handlar om att göra och utvärdera ett försök från en kvantitativ analys. Det skall vara en hjälp för dig som inom ditt kandidat- eller mastersarbete¹ skall samla in observationer som ger ett siffermaterial.

Det du vill studera kan vara höjden av plantor, vikten av biomassa, etc., men kan också vara antalet plantor som kommer upp eller en bedömning av plantornas kvalité om denna anges på en numerisk skala. Många av råden i kompendiet går också att använda vid enkätstudier, även om enkätstudier ofta är observationsstudier² där du inte kan styra försöksupplägget på samma sätt som vid de experiment som finns i kompendiets exempel.

Kompendiet behandlar inte arbeten som är litteraturstudier även om vissa avsnitt kan hjälpa dig att förstå och kritiskt granska vetenskapliga artiklar.

1.2 Arbetsgång för examensarbetet

Det kan vara handledaren som föreslår vad det är för problem du skall undersöka, men bortsett från det så beskriver Figur 1.1 arbetsgången i ditt examensarbete. Den sista pilen visar att ditt resultat kan leda till nya idéer för senare examensarbeten.



Figur 1.1.

Arbetsgång för ett examensarbete baserat på ett experiment.

¹ Kandidat- och mastersarbeten kallas i fortsättningen gemensamt för examensarbeten.

² Skillnaden mellan observationsstudie och experiment förklaras i Avsnitt 1.4.

1.3 Vetenskap och beprövad erfarenhet

Inom medicinen finns kravet att de behandlingar som utförs skall vara baserade på vetenskap och beprövad erfarenhet. Vid ett examensarbete är det svårt att basera resultaten på beprövad erfarenhet men resultatet bör, så långt det är möjligt, byggas på vetenskapligt förankrade slutsatser.

Det innebär att det finns någon form av garanti för att resultatet som du redovisar från försöket inte är en slumpmässig effekt utan att det beror på en verklig skillnad mellan de olika behandlingarna. Kan du vara säker på att resultatet blir detsamma om försöket upprepas under samma förutsättningar?

Tiden och resurserna räcker inte alltid till för att inom ett examensarbete göra så stora försök som behövs för att kunna visa att det verkligen är en skillnad mellan behandlingarna. Det är då viktigt att du är medveten om detta och tydligt skriver att den skillnad som finns i försöket inte är statistiskt signifikant³ men att resultatet kan ligga till grund för framtida försök.

Detta kompendium börjar från början och beskriver arbetet från hypoteser och försöksdesign, via inläsning av data i datorn, till statistiska beräkningar och vetenskapligt förankrade slutsatser. Du behöver kanske inte alla delarna men då är det bara att gå till det avsnitt som är aktuellt för dig.

1.4 Experiment eller observationsstudie

Variabler är ett viktigt begrepp i statistiken, men det är lite förvirrande att variabler i statistiken inte har riktigt samma betydelse som i matematiken.

En *variabel* i statistiken är någon av de faktorer som finns med i försöket. Det kan vara bevattning, gödselmängd, ljus, skörd, planthöjd, etc.

Några av dessa variabler kan du styra själv, t ex bevattning, ljus och gödsling, och även om de är konstanta genom hela försöket så måste du anteckna vad de har för värde eftersom beskrivningen i ditt examensarbete skall göra försöket *replikerbart*. Det betyder att din beskrivning av hur försöket har gått till skall vara så noggrann att en annan person skall kunna göra försöket exakt så som du gjorde det för att se om dina slutsatser är rimliga⁴.

Du skall därför genom hela arbetet anteckna vad du gjorde och när du gjorde det, även om du just då inte tycker att det är någon intressant information. Tro inte att du i efterhand kommer ihåg exakt när och hur mycket du vattnade om du inte har skrivit ner det!

³ Signifikant kommer att vara ett vanligt ord i det här kompendiet, den som vill kan ersätta det med "tydlig" som talar om vad det handlar om.

⁴ Beskrivningen av försöket skall helst också göra det *reproducerbart*. Detta innebär att det skall gå att göra om de beräkningar som du har gjort utifrån beskrivningen av statistiska metoder etc. som finns i examensarbetet.

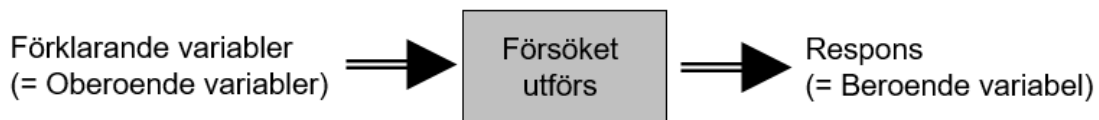
Det finns tre typer av variabler i försöket:

- Variabler som inte direkt har med försöket att göra men som ändå kan noteras (temperatur, tid på året, försöksplats, pH-värde i jorden)⁵.
- Variabler som du styr för att se om de har effekt på resultatet (bevattning, gödning, ljus).
- Variabler som är ett resultat av försöket⁶ (skörd, planthöjd, plantdiameter).

Det som utmärker ett *experiment* (detta kallas också ofta ett *försök*) är att du själv kan påverka många av variablerna. De variabler som du kan styra kallas *förklarande variabler* (eller oberoende variabler) och de variabler som är ett resultat av försöket och som du inte kan styra kallas *responser* (eller beroende variabler).

Bevattning och jordbearbetning är exempel på förklarande variabler som du kan påverka, medan avkastningen är respons i försöket. Det som gör det till ett experiment är att bara de förklarande variablerna ändras och annat hålls på en så konstant nivå som möjligt. Om det då blir en tydlig skillnad i responsen för de olika behandlingarna kan du vara ganska säker på att det beror på de förklarande variabler som du har ändrat på.

Är du bara intresserad av att se en effekt av bevattningen så skall du inte samtidigt ändra på jordbearbetningen eftersom du då inte vet om det är bevattningen eller jordbearbetningen som har orsakat den observerade skillnaden.



Figur 1.2.

Schematisk beskrivning av ett experiment.

När det bara är observationer utan att du kan styra några variabler, t ex då du skickar ut en enkät till slumpmässigt valda personer, så kallas det en *observationsstudie*. Eftersom du inte har kunnat styra ålder, bostadsort, etc. i utskicket så kan du inte vara helt säker på vad det är som har gjort äldre och yngre har olika åsikt. Det behöver ju inte bara bero på åldern, det kan ju lika gärna bero på att de äldre har högre lön, fast jobb, etc.

Det är viktigt att notera om det är en observationsstudie eller ett experiment när du tolkar resultatet från olika studier. Ett experiment är mer pålitligt eftersom du själv har kunnat ge så lika förutsättningar som möjligt för allt som inte ingår i problemställningen.

Nackdelen med ett experiment är att det kan vara svårt att generalisera när dina resultat är gjorda under speciella förutsättningar. Denna fråga diskuteras ganska

⁵ Om det sedan visar sig att de har påverkat försöket så brukar dessa variabler kallas ibland för kovariater.

⁶ Eftersom resultatet av dessa variabler delvis beror på slumpen så kallas de ibland för slumpvariabler.

ofta i medicinska studier; får man en effekt av en behandling på studenter som frivilligt har ställt upp i en studie så är det inte säkert att det går att generalisera och säga att det blir en effekt av behandlingen för alla människor.

Här beskrivs experiment (eller försök) där responsen *alltid* är ett numeriskt värde (= kvantitativ) medan den förklarande variabeln kan vara text (= kvalitativ), t ex kan de grupper du jämför vara "behandlat" eller "obehandlat".

Försöksplanen skall så noggrant som möjligt visa hur försöket ser ut, men under och efter försöket bör du också notera saker som inte gick att påverka i försöksplanen, t ex temperatur, nederbörd, etc.

1.5 Statistiska hypoteser

En grundläggande idé i den vetenskapsfilosofi som hypotestesten bygger på är att hypotesen skall kunna motbevisas⁷. För att du skall kunna motbevisa en hypotes så formuleras de statistiska hypoteserna lite annorlunda än i det vanliga språket. Exempelvis kan hypoteserna se ut så här:

På vanlig svenska är hypotesen:

"Behandlat och obehandlat ger olika skörd."

De statistiska hypoteserna:

Nollhypotes: "Behandlat och obehandlat ger samma skörd."

Mothypotes: "Behandlat och obehandlat ger olika skörd."

När du har samlat på dig tillräckligt mycket argument för att det är en skillnad så kan du säga att nollhypotesen förmodligen inte är sann. Du har då motbevisat nollhypotesen och det uttrycks i statistiken som att du *förkastar* nollhypotesen till förmån för mothypotesen.

Så länge du inte har övertygande bevis för att säga att nollhypotesen inte är sann så får du fortsätta tro på den och detta formuleras i statistiken som att du *inte förkastar* nollhypotesen.

1.6 Replikat

För att förkasta nollhypotesen räcker det inte att hitta *ett* fall som talar mot nollhypotesen eftersom det också finns en slumpmässig variation. I alla kurser i statistik poängteras hur viktigt det är med replikat (= upprepningar) för att kunna uppskatta variationen. En observation av varje behandling räcker inte och det framgår av detta exempel som visar hur fel det kan bli med bara en observation av varje behandling.

⁷ Ett liknande uttryck för detta är att hypotesen skall kunna falsifieras.

Exempel 1.1.

Det finns två försöksrutor⁸ placerade intill varandra och dessa kallas för Ruta 1 och Ruta 2. Om båda rutorna är obehandlade så blir skördarna i försöksrutorna

Ruta 1: (obehandlad) 14,2 <i>kg</i>	Ruta 2: (obehandlad) 13,0 <i>kg</i>
---	---

Eftersom båda rutorna är behandlade på exakt samma sätt (dvs. obehandlade) kan du förvänta dig lika stora skördar. Så blir det som regel inte på grund av ojämnheter i jorden, packning, gödselspridning, mätfel etc.

I detta fall är skillnaden 1,2 *kg*. Skillnaden är stor men inte extrem.

Nu har en av rutorna fått en behandling av något slag, t ex ett nytt gödningsmedel, och denna behandling har resulterat i en faktisk skördeökning på 1,0 *kg* per ruta.

Om Ruta 1 är behandlad och Ruta 2 är obehandlad blir försöksresultatet

Ruta 1: (behandlad) 15,2 <i>kg</i>	Ruta 2: (obehandlad) 13,0 <i>kg</i>
--	---

Här blir alltså resultatet att behandlingen är en framgång och ger en skördeökning med 2,2 *kg* per ruta.

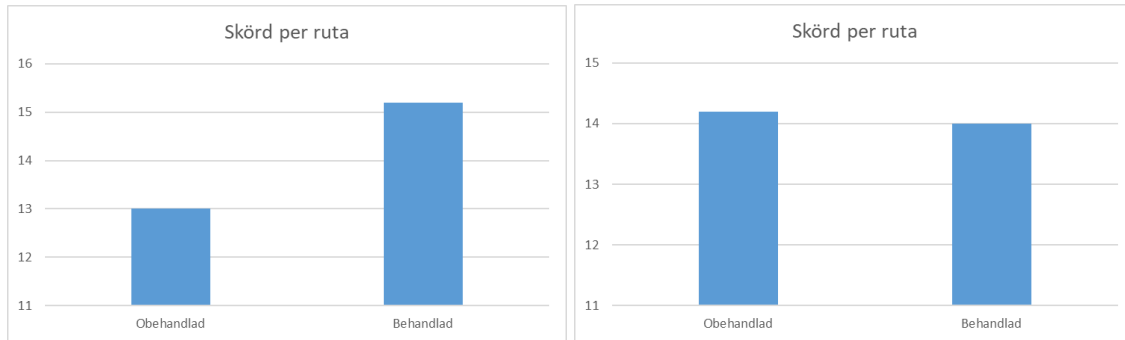
Om istället Ruta 2 är behandlad och Ruta 1 är obehandlad blir försöksresultatet

Ruta 1: (obehandlad) 14,2 <i>kg</i>	Ruta 2: (behandlad) 14,0 <i>kg</i>
---	--

I detta fall har försöket felaktigt gett resultatet att behandlat är sämre än obehandlat och ger en skördeminskning med 0,2 *kg* per ruta. Detta beror på att den slumpmässiga variationen i det här fallet är större än behandlingseffekten. Om du helt och hållet litar på dessa två observationer så har du fått en motsägelse mot påståendet att behandlingen ger en skördeökning, men det beror på att den slumpmässiga variationen är större än behandlingseffekten.

Genom att slumpmässigt välja vilken ruta som skall behandlas så går det att uttala sig lite ärligare efter försöket och det går att bemöta kritiken att behandlingen medvetet lades på den ruta som var bäst. Men det räcker inte, det behövs också replikat för att kunna uppskatta slumpvariationen, dvs. hur mycket skörden kan variera för de rutor som har fått samma behandling.

⁸ Försöksrutor kallas ibland för parceller.



Figur 1.3.

Resultaten inlagda med standardinställningar för stapeldiagram i Excel. Kalkylprogrammet har "huggit av" axlarna så att skillnaden ser större ut än vad den i själva verket är. Skalorna på axlarna är inte heller samma i båda figurerna, vilket också är viktigt att lägga märke till när man jämför olika försök.



Det går att dra några viktiga slutsatser av detta:

- Ett enskilt försök utan replikat, dvs. endast en observation av varje behandling, kan ge vilseledande resultat om den slumpmässiga variationen är större än behandlingseffekten.
- Den statistiska analysen går ut på att försöka skilja på behandlingseffekten och slumpvariationen för att i bästa fall se att behandlingseffekten är dominerande. Utan replikat går det inte att uppskatta och skilja den slumpmässiga variationen (som innefattar hur ojämn försöksplatsen är, variation mellan plantor, etc.) från behandlingseffekten. Med fler replikat blir också den slumpmässiga variationen mindre.
- Trots bristerna kan det ibland finnas skäl att arbeta med försök utan replikat, men knappast i ett examensarbete. Odlaren vill kanske göra egna små försök och har kanske från tidigare en uppfattning om hur mycket skörden varierar. Det är då viktigt att vara medveten om att resultaten måste tolkas med största försiktighet.
- ***Var alltid kritisk när du tolkar resultaten av ett försök!***

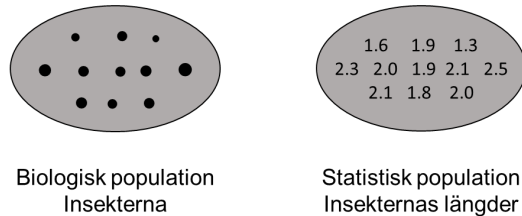
1.7 Population

Hypoteserna handlar oftast om att jämföra olika populationer, och därför är det bra att redan från början veta vad som skiljer en statistisk population från en biologisk population. I beskrivningen i det här avsnittet är det bara *en* population men hypotesen handlar oftast om att jämföra två eller flera populationer som har fått olika behandlingar.

En statistisk *population* (eng. population) är mer abstrakt än en biologisk population. Skillnaden mellan en biologisk population och en statistisk population framgår av ett exempel.

Exempel 1.2.

Du vill bestämma längden på en insekt av en viss typ. Den biologiska populationen är alla de insekter som kan förekomma, men den statistiska populationen är *längderna* av alla insekterna. Den statistiska populationen är alltså en massa siffervärden (lika många som antalet insekter i den biologiska populationen).

**Figur 1.4.**

En biologisk population och motsvarande statistiska population. ■

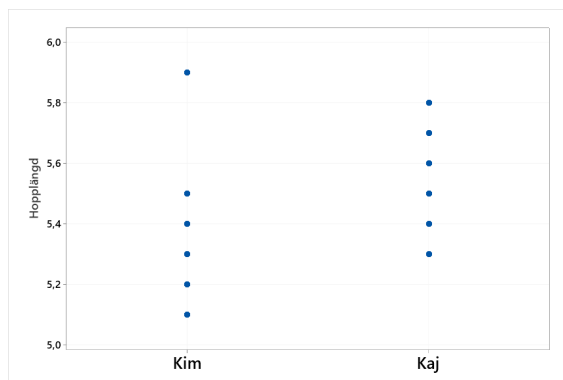
Vad menas då med att bestämma insekternas längder i populationen? Eftersom alla insekter inte är lika långa räcker det inte att mäta en insekt och ange det som ett mått på längden. Insekternas längd är ett exempel på en *slumpvariabel* (eng. random variable) eftersom längden varierar beroende på vilken insekt du väljer.

Teoretiskt skulle du kunna mäta längden på alla insekter i populationen och illustrera det med ett histogram, men det är i praktiken omöjligt. Det är också ett onödigt arbete och en onödig kostnad eftersom din hypotes oftast bara gäller populationens medelvärde.

Exempel 1.3.

Det är oftast medelvärden som är intressanta i statistiken och här skiljer sig statistiken helt från en idrottstävling.

I en traditionell längdhoppstävling vinner den som hoppar längst på sex försök medan vinnaren i statistiken är den som har *högst medelvärde* i det långa loppet. I OS eller VM skall man ha bästa tiden eller längsta hoppet, men i statistiken gäller det istället att vara bäst "i medeltal", och detta är illustrerat i Figur 1.5.

**Figur 1.5.**

Längdhopsresultat från Kim och Kaj där Kaj är bättre i statistisk mening trots att Kim har det längsta hoppet.

Kim har medelvärdet 5,40 meter och Kaj har medelvärdet 5,55 meter och därför anses Kaj vara bättre i statistisk mening även om Kim skulle vinna tävlingen med det längsta hoppet på 5,90 meter. ■

Populationer kan vara olika stora. Skörden per hektar i korn är ett exempel på en population som teoretiskt är oändlig eftersom det i teorin går att göra detta försök på hur många platser och under hur många år som helst. För att studera åldern hos de människor som hittills har varit på till månen handlar det om en mycket liten och ändlig population.

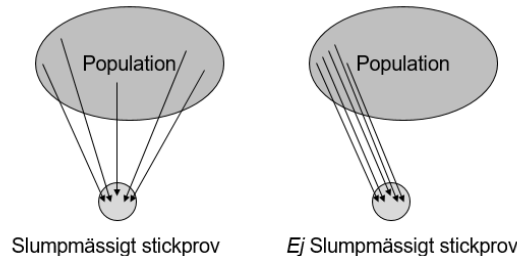
I detta kompendium antar vi oftast att populationen är oändlig, eller åtminstone så stor att den kan betraktas som oändlig⁹.

⁹ När man gör olika typer av enkätundersökningar för det svenska folket räknar man oftast som att populationen är oändlig även om den i praktiken är ändlig.

2 Försöksplanering

2.1 Slumpmässigt stickprov

För att få en bild av populationen tar du ett stickprov (eng. sample) och hoppas att detta stickprov är likt populationen.

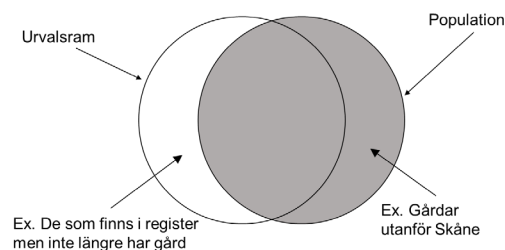


Figur 2.1.

Ett slumpmässigt stickprov och ett stickprov som inte är slumpmässigt.

För att du skall kunna använda stickprovet för att säga något om hur populationen ser ut skall det vara ett *slumpmässigt stickprov*. Detta innebär att alla element i populationen har samma chans att komma med i stickprovet *oberoende av varandra*. Du kommer att upptäcka att det är ganska svårt att göra ett perfekt slumpmässigt stickprov i praktiken men det är viktigt att få det så korrekt som möjligt eftersom du inte kan rätta upp detta i efterhand.

- För att säga något om storleken på morötter kan du inte att välja en gård slumpmässigt och låta det slumpmässiga stickprovet vara alla morötter som skördas på den gården. Även om gården är slumpmässigt vald är inte kravet på oberoende uppfyllt; om en morot från gården är vald så kommer de andra också att väljas ut. Är detta en gård som generellt har stora morötter så blir den en skev bild av populationen av morötter. Detta brukar även på svenska beskrivas som att stickprovet är *biased*.
- För att uttala sig om gårdar i Sverige går det inte att välja gårdarna slumpmässigt från en lista över gårdar i Skåne. I enkätundersökningar brukar detta beskrivas som att urvalsramen inte är densamma som den population man är intresserad av.



Figur 2.2.

Urvalsram och population för gårdar i Sverige i ett urval från Skåne.

2.2 Randomisering

Stickprovet skall alltså tas slumpmässigt (randomiserat, eng. randomised) för att det skall gå att använda i den statistiska analysen. Detta innebär att alla enheter skall ha samma möjlighet att komma med oberoende av vilka andra enheter som har valts ut. Efter att ha numrerat alla enheterna finns det några olika sätt att få slumpmässiga stickprov, t ex

- lägg lappar i en skål och dra lapparna slumpmässigt en och en¹⁰,
- använd slumptalsfunktionen i Excel eller ett statistiskt programpaket.

Det är självklart hur du drar lappar ur en skål eller en hatt, men det är inte lika lätt med slumptalsfunktionen i Excel och Exempel 2.1 visar därför hur detta går till.

Exempel 2.1

Du har fått 16 plantor för ditt försök men får reda på att det bara finns plats för åtta plantor i växthuset i den första försöksomgången. Du skall därför slumpmässigt välja ut åtta av plantorna till denna första försöksomgång.

Du numrerar därför plantorna från 1 till 16 och använder Excel för att bestämma vilka åtta plantor som skall vara med i försöket.

- Skriv siffrorna 1 till 16 i den första kolumnen, där varje siffra anger en planta.
- I kolumnen bredvid skrivs =SLUMP() (parentesen måste vara med) i de 16 första raderna.
- Markera sedan de två kolumnerna, välj fliken Data, därefter Sortera och välj att sortera efter den kolumn som innehåller slumptalen.
- Nu finns siffrorna 1 till 16 i slumpmässig ordning i den första kolumnen och du väljer de åtta första till ditt försök.
- Varje gång du rör något i kalkylbladet så ändrar sig slumptalen utan att de sorteras om och därför syns det inte att det är ordnat från minsta till högsta värde; efter sorteringen kom det nya slumptal!

	A	B	C	D	E
1	7	0.299004			
2	1	0.018053			
3	16	0.980259			
4	6	0.456819			
5	9	0.799603			
6	10	0.519145			
7	11	0.053623			
8	3	0.957379			
9	5	0.309367			
10	2	0.148912			
11	4	0.297734			
12	14	0.065998			
13	15	0.339376			
14	13	0.977982			
15	8	0.219271			
16	12	0.276614			
17					

¹⁰ Eftersom man i det här fallet inte vill ha samma siffra flera gånger så drar man utan att lägga tillbaka den dragna lappen ur hatten.

- Du hade självklart kunnat göra samma sak med lappar och om du fått lapparna med nummer 7, 1, 16, 6, 9, 10, 11, 3 så hade det blivit samma resultat av randomiseringen. I detta fall är det ju inte någon stor tidsvinst att använda datorn. ■

2.3 Frågeställningar

För att du skall kunna göra en försöksplan måste du först ha klart för dig hur ditt försök ser ut. Det är framförallt tre frågor som du skall ta ställning till innan du gör försöksplanen.

- Vad är experimentenheten i det här försöket?
- Hur många olika behandlingar finns det i försöket?
- Finns det någon blockeffekt att ta hänsyn till?

Experimentenheten definieras som den minsta enhet som kan få en egen behandling. Om det är ett växthusförsök med bevattning och krukor så är ofta krukans experimentenhet eftersom två plantor i samma kruka måste få samma bevattning. Experimentenheten är kopplad till att det skall vara oberoende observationer, om du gör flera mätningar på samma experimentenhet så är resultatet av dessa mätningar förmodligen beroende av varandra.

Den andra frågan är lättare att svara på och det enda du skall avgöra är om det är två behandlingar eller mer än två behandlingar. Två behandlingar är t ex när du jämför behandlat mot obehandlat, men i de allra flesta fall är det mer än två behandlingar i försöket, t ex om du har flera olika gödningspreparat.

Den tredje frågan om block återkommer senare i kompendiet, men block innebär att det finns block där förutsättningarna inom blocket är ganska lika. När det är två behandlingar skulle ett block¹¹ kunna bestå i att tvillingar sätts ihop i block och den ena tvillingen får en behandling och den andra tvillingen får den andra behandlingen. Eftersom de är lika i övrigt så kan man lättare se skillnaden i behandlingen.

För mer än två behandlingar så kan blocket bestå av odlingsborden i växthuset. Ett av borden kanske står i solen, ett annat står så att det får mycket vatten, etc. På varje odlingsbord finns *alla* behandlingar och det gör att det går att jämföra behandlingarna inom samma bord.

Med denna bakgrund går vi nu vidare och ger exempel på hur man randomiserar i de fyra fall som finns ovan:

- Två behandlingar utan block.
- Två behandlingar med block (= par).
- Mer än två behandlingar utan block.
- Mer än två behandlingar med block.

Har du redan ett arbete aktuellt så fundera gärna på vilken kategori ditt experiment tillhör.

¹¹ När det bara är två behandlingar kallas det ofta ett par.

2.4 Två behandlingar

När du har två behandlingar och gör en randomisering så kan man kalla försöket ett experiment. Det som gör det till ett experiment är att du på förhand genom randomisering bestämmer vilken ruta eller planta som skall ha vilken behandling och förutom de olika behandlingarna så skall rutorna helst vara så lika som möjligt.

Försöksrutorna kan ju inte vara helt lika och för att så långt som möjligt ta bort effekten av att de skiljer sig åt så väljer du slumpmässigt ut vilken ruta som får vilken behandling. Då kan du inte i efterhand få kritik för att ha valt de bästa rutorna till den behandling som du vill skall vara bäst.

Med två behandlingar (varav en möjligen är en kontroll eller obehandlad) så är det nästan alltid bäst att ha lika många rutor av varje behandling, och detta beskrivs som att försöket är *balanserat*. I detta exempel kallas behandlingarna för A respektive B (där eventuellt A kan vara behandlingen "obehandlad").

Exempel 2.2

Det finns tolv försöksrutor och du skall slumpmässigt fördela så att sex rutor behandlas med A och sex rutor behandlas med B. Du börjar med att numrera rutorna från 1 till 12:

1	2	3	4	5	6	7	8	9	10	11	12
---	---	---	---	---	---	---	---	---	----	----	----

Om lapparna eller Excel bland 1 till 12 väljer ut siffrorna 4, 11, 6, 1, 10, 8 och du bestämmer att de sex utvalda rutorna får behandling A så blir försöksplanen

A	B	B	A	B	A	B	A	B	A	A	B
---	---	---	---	---	---	---	---	---	---	---	---

Eftersom du har slumpat ut rutorna som får behandling A utan några som helst begränsningar mer än att det skall vara sex rutor av varje behandling så beskriver du detta som ett *fullständigt randomiserat försök*. När du sedan har samlat in värdena och skall analysera dem så hittar du metoderna i Avsnitt 5.5. ■

Det är självklart att du skall randomisera försöket för att inte få kritik, men det kan ju inträffa att du har otur i randomiseringen och får försöksplanen

A	A	A	A	A	A	B	B	B	B	B	B
---	---	---	---	---	---	---	---	---	---	---	---

Accepterar du den här randomiseringen? Det lättaste är förmodligen att göra om randomiseringen även om det inte är helt tillåtet, men ännu bättre är att fundera på varför du inte tycker om den här försöksplanen. Det beror förmodligen på att du misstänker att skörden kan variera i väst-östlig riktning. Det går därför inte att veta om en observerad skillnad mellan A och B beror på att det verkligen är en skillnad eller på att fältet har en variation i väst-östlig riktning.

Denna typ av problem bör du ha insett redan före randomiseringen, och då skulle du ha övervägt om det inte är bättre att ersätta det fullständigt randomiserade försöket med ett blockförsök. Ett randomiserat *blockförsök* med två behandlingar kallas också *stickprov i par* eller matchning eftersom det är ett par av rutor i varje block.

Exempel 2.3

i detta fall delar du in fältet i sex block (block betecknas ofta med romerska siffror) om vardera två intilliggande rutor och låter slumpen avgöra vilken av de två rutorna som får behandling A respektive B.

Block I		Block II		Block III		Block IV		Block V		Block VI	
Ruta	Ruta	Ruta	Ruta	Ruta	Ruta	Ruta	Ruta	Ruta	Ruta	Ruta	Ruta
1	2	3	4	5	6	7	8	9	10	11	12

Vid randomiseringen så går det att använda samma metod som i ett fullständigt randomiserat försök ovan och numrera rutorna men då *inom varje block*. Med bara två alternativ kan du singla en enkrona och säga att om det blir krona så väljs A i den första rutan. Eftersom det skall vara en behandling av varje i alla block så blir det därför B i den andra rutan, medan om myntet visar klave så väljs B i den första rutan och A i den andra rutan. Med myntkasterna Krona, Krona, Klave, Klave, Klave, Krona så blir försöksplanen

I		II		III		IV		V		VI	
A	B	A	B	B	A	B	A	B	A	A	B

Här är två behandlingar, men analysen beräknar differensen mellan A och B inom varje block. I Avsnitt 5.4 beskrivs hur du analyserar försök som är upplagda på detta sätt. ■

2.5 Mer än två behandlingar

Med mer än två behandlingar analyseras värdena med variansanalys. Här finns också alternativet att ha ett fullständigt randomiserat försök samt att ha ett försök med block.

Exempel 2.4

I ett *fullständigt randomiserat försök* med tre behandlingar A, B och C så skall du lägga ut 4 rutor av varje behandling om det är ett balanserat försök. Det går att göra randomiseringen på lite olika sätt, förmodligen är det också här lättast att göra 12 lappar och dra dessa slumpmässigt. Numrera de 12 rutorna från 1 till 12 på samma sätt som tidigare och låt de fyra först valda få behandling A. Detta blir då rutorna 4, 11, 6 och 1. Sedan går du vidare och tittar vilka rutor som skall vara till B, och detta blir då rutorna 10, 8 (fortsätt att dra lappar), 9 och 12. Självklart blir det C i de fyra rutor som är kvar och försöksplanen blir därför

A	C	C	A	C	A	C	B	B	B	A	B
---	---	---	---	---	---	---	---	---	---	---	---

När du har lagt ut denna typ av försök och har samlat in observationerna så skall du göra den statistiska analysen med variansanalys som beskrivs i Avsnitt 5.8. ■

Någon kan ju här invända att om det är en variation i skörd i väst-östlig riktning så kommer B att få bättre/sämre förutsättningar än de andra när alla rutor med B ligger i öster. Om resultatet visar att B är bättre än de andra behandlingarna så är det då svårt att veta om detta beror på en verklig behandlingseffekt eller på att fältet är bättre i de östra delarna.

För att gardera sig mot denna typ av kritik typ så kan man göra ett *blockförsök* med fyra block.

Exempel 2.5

De fyra blocken läggs ut så att rutorna inom blocken är så lika som möjligt samt att varje block innehåller varje behandling precis en gång. Utläggningen av behandlingar inom block gör du som tidigare med randomisering. Med fyra block blir försöksplanen

I			II			III			IV		
Ruta 1	Ruta 2	Ruta 3	Ruta 4	Ruta 5	Ruta 6	Ruta 7	Ruta 8	Ruta 9	Ruta 10	Ruta 11	Ruta 12

Randomiseringen kan ske på olika sätt. Det minst komplicerade är att göra en skål med tre lappar märkta A, B och C. Sedan drar du lapparna en och en och fyller i det första blocket. Det är ju onödigt att dra den tredje lappen eftersom det är klart vilken den är. Lägg tillbaka lapparna i hatten och dra igen, etc.

Med resultatet B, A (och självklart C i sista rutan i Block I) blir det B A C i första blocket. Fortsätter man på samma sätt med de övriga blocken får man t ex

Block II: A, B (och C skrivs i sista rutan i Block II).

Block III: A, C (och B skrivs i sista rutan i Block III).

Block IV: B, C (och A skrivs i sista rutan i Block IV).

Sammanfattningsvis har försöksplanen i det här fallet blivit

I			II			III			IV		
B	A	C	A	B	C	A	C	B	B	C	A

Lägg märke till att du gör en ny randomisering för varje block men att det inte är något som hindrar att randomiseringen ger två block som är lika.

När du har lagt ut denna typ av försök och har samlat in observationerna så skall du göra den statistiska analysen med hjälp av variansanalys med block som beskrivs i Avsnitt 5.9. ■

3 Insamling av data

3.1 Generella råd

Planera redan från början hur du skall göra ditt försöksprotokoll i Excel där du så småningom skall fylla i resultatet av försöket. Titta då i Avsnitt 3.3 så att du vet hur du skall läsa in dina data. Även om du inte tänker använda Excel för att göra beräkningarna så blir det överskådligt med Excel och det är väldigt lätt att flytta materialet till andra program om du vill det.

Några generella råd redan från början:

- Gör ett kalkylblad som *bara* innehåller rådata.
- Kopiera sedan kalkylbladet med rådata till blad där du kan göra beräkningar och figurer.
- Det är bra att dela upp beräkningarna på många kalkylblad så det är lätt att hitta.
- Ge bladen namn så att du hittar dina beräkningar och figurer.
- Lägg märke till att Excel eller andra datorprogram kan göra alla beräkningarna, *du behöver inte använda miniräknare!*

3.2 Olika typer av variabler

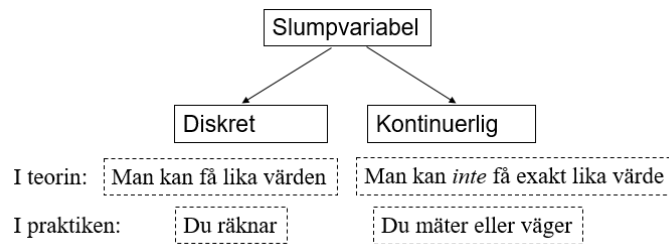
Kompendiets begränsning till kvantitativa observationer gäller självklart bara resultatet av försöket. De förklarande variablerna i experimentet behöver inte vara kvantitativa och kan beskrivas med ord. I Kapitel 2 har ju behandlingarna och blocken varit kvalitativa och beskrivits med A, B och C respektive I, II, III och IV.

I det här avsnittet diskuterar vi hur man beskriver resultatet av försöket och det är då en kvantitativ variabel.

När du illustrerar resultatet av ett stickprov från ett kvantitativt försök så går det antingen att göra ett stolpdiagram eller ett histogram och detta delar också variablerna i ditt resultat i två olika kategorier.

Antalet blommor på en planta är ett exempel på en *diskret slumpvariabel* (eng. discrete random variable). Det är något du räknar och kan bara anta heltalsvärden, 3, 7, 63, osv. Det går aldrig att få 12,5 blommor på en planta. Du registrerar den här typen av variabler som heltal, men du kan senare få decimaler när du räknar ut medelvärdet eller standardavvikelsen av antalet blommor på ett antal plantor.

Skörden i *kg* är ett exempel på en *kontinuerlig slumpvariabel* (eng. continuous random variable). Det är något som du mäter eller väger och kan inom vissa gränser anta alla tänkbara värden, t ex 120,1142739 *kg*. Det går att fortsätta med hur många decimaler som helst och i teorin finns det ett oändligt antal decimaler men mätapparaturen sätter en gräns för noggrannheten. När du registrerar och gör beräkningar på den här typen av variabler är det självklart bra att ha med så många decimaler som möjligt.

**Figur 3.1.**

Diskret respektive kontinuerlig slumpvariabel.

En diskret slumpvariabel illustreras bäst med ett stolpdiagram¹² medan en kontinuerlig slumpvariabel illustreras med ett histogram där datorn klassindelar data-materialet. Skillnaden är alltså att stolpdiagrammet inte kan få värden mellan stolparna, medan histogrammet kan få värden var som helst och staplarna kan göras smalare om det kommer in fler observationer.

I Kapitel 4 finns beskrivningar av hur du gör dessa och andra typer av diagram i Excel.

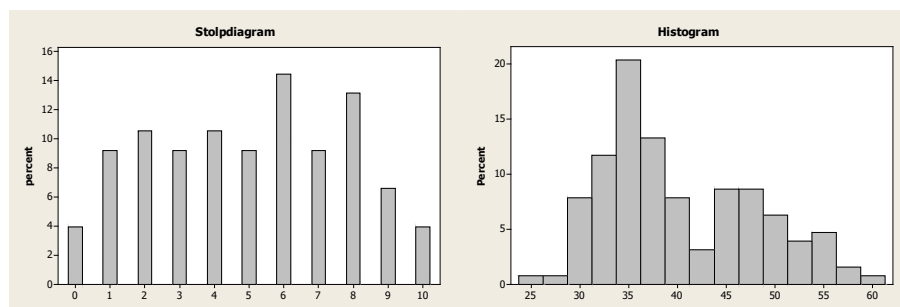
**Figur 3.2.**

Illustration av diskret respektive kontinuerlig slumpvariabel¹³.

I figurerna är enheten procent på andra axeln och detta kallas ibland för *relativ frekvens*. Relativ frekvens har fördelen att summan av stolparnas höjder i ett stolpdiagram är lika med 1 (eller 100 %) och ytan i ett histogram är 1 (eller 100 %). Detta betyder också att diagrammen med relativ frekvens inte växer över alla gränser när det blir fler observationer och att enheten på andra axeln aldrig kan gå över 100 %. I ett stolpdiagram med relativ frekvens är stolpens höjd vid värdet 0 (t ex) lika med

$$\text{"relativ frekvens av 0"} = \frac{\text{"antal observationer som är 0"}}{\text{"totala antalet observationer"}}$$

¹² Skillnaden mellan stolpdiagram och stapeldiagram är lite oklar. Det brukar beskrivas som att ett stolpdiagram alltid har siffervärden på den första axeln under stolparna medan ett stapeldiagram kan ha olika färger, olika bilmärken eller andra icke-numeriska på den första axeln under staplarna.

¹³ Dessa och många andra figurer i kompendiet är gjorda i det statistiska programpaketet Minitab.

3.3 Observationer i datorn

Då du har samlat in observationerna och skall använda en dator för att analysera datamaterialet bör detta läggas in på ett sätt som gör att det fortsatta arbetet blir så enkelt som möjligt. Trots att det finns olika datorprogram för statistisk bearbetning gäller generella regler för hur datamaterialet skall struktureras.

Materialet läses in i ett rektangulärt schema, där raderna ("horisontellt") är observationer och kolumnerna ("vertikalt") är variablerna, dvs. de förklarande variablerna och de responsvariabler som man mäter. Gör man så här är det lätt att lägga till nya observationer efterhand som de kommer in.

	Variabel 1	Variabel 2	Variabel 3	...
	↓	↓	↓	
Observation 1 →				...
Observation 2 →				...
Observation 3 →				...
	⋮	⋮	⋮	

Exempel 3.1.

Vid ett växthusförsök för att undersöka bevattningens effekt på skörden och smaken i en skala från 1 (dålig) till 5 (bra) delades växthusets fyra bord i fyra block som betecknades I, II, III och IV. De olika behandlingarna kallades för A, B och C, där

- A = halv dos mot rekommenderad
- B = rekommenderad dos
- C = dubbel dos mot rekommenderad

Datamaterialet kan beskrivas med tolv rader och fyra kolumner med en extra första rad som anger variablernas namn. Här är Block och Behandling förklarande variabler, Skörd en kontinuerlig responsvariabel och Smak en diskret responsvariabel.

Block	Behandling	Skörd	Smak
I	A	223,5	2
I	B	242,1	5
I	C	168,6	4
II	A	246,2	
II	B	229,1	4
II	C	178,1	3
III	A	269,0	3
III	B	208,4	
III	C	281,8	5
IV	A	259,4	1
IV	B	282,3	4
IV	C	192,7	2

Smaktesterna gjordes av en och samma person för att kunna jämföra på ett korrekt sätt. I tabellen finns två saknade värden i kolumnen för smak, och saknade värden kan ibland orsaka problem eftersom det har olika beteckning i olika program. Excel klarar av en blank cell och använder inte cellen vid beräkning av medelvärden etc. I Minitab används asterisk (*) för saknat värde, i SAS är det punkt (.) och i R är det NA.

I Excel hade du alltså lagt in det i ett kalkylblad med (t ex) namnet Rådata.

	A	B	C	D	E
1	Block	Behandling	Skörd	Smak	
2	I	A	223,5	2	
3	I	B	242,1	5	
4	I	C	168,6	4	
5	II	A	246,2		
6	II	B	229,1	4	
7	II	C	178,1	3	
8	III	A	269	3	
9	III	B	208,4		
10	III	C	281,8	5	
11	IV	A	259,4	1	
12	IV	B	282,3	4	
13	IV	C	192,7	2	
14					

De celler som Excel uppfattar som numeriska värden står till höger i cellen och de som uppfattas som text är till vänster; detta är en bra metod för att se om Excel använder decimalkomma eller decimalpunkt. I detta fall är det decimalkomma eftersom skördarna står till höger i cellen. ■

3.4 Några vanliga frågor

Decimalkomma eller decimalpunkt?

I svensk litteratur används ofta decimalkomma och så är det i detta kompendium, men i engelska program är det ofta decimalpunkt. Det kan orsaka problem när du flyttar ditt datamaterial mellan olika program, men i moderna program läser ofta programmet av detta automatiskt. Det är bra att kontrollera om programmet är inställt på decimalpunkt eller decimalkomma från början. *Några figurer i det här kompendiet har decimalpunkt eftersom det var inställningen i det program där figurerna gjordes.*

Hur många decimaler?

Att skriva t ex 8 kg innebär att resultatet angetts med en siffras noggrannhet. Ur statistisk synpunkt innebär det att värdet kan ligga mellan 7.5 och 8.5 kg. Skriver vi 8.0 kg innebär det att värdet kan ligga mellan 7.95 och 8.05 kg. Tre siffrors noggrannhet innebär på motsvarande sätt att 8.34 kg medför att värdet kan ligga mellan 8.335 och 8.345 kg. Man skall inte ange sitt värde med större noggrannhet än mätmetoden tillåter. Samma gäller resultatet efter bearbetning och analys. *Däremot skall man under räknearbetets gång arbeta med så många siffror som möjligt.*

4 Analys av data

4.1 Inledning

För att illustrera försöksresultatet så används tabeller eller figurer. I det här kapitlet handlar de första avsnitten om de vanligaste sammanfattande numeriska måtten som finns i tabellerna och de senare avsnitten handlar om hur du gör för att illustrera med figurer.

4.2 Medelvärde

Det vanligaste lägesmättet är det *aritmetiska medelvärdet* (eng. arithmetic mean eller average), som oftast kallas enbart medelvärdet¹⁴. Eftersom datorn automatiskt räknar ut medelvärdet behöver du inte bry dig om exakt hur formeln ser ut, men för fullständighets skull skriver vi ändå ner formeln här. Det viktiga är att veta att den generella beteckningen för medelvärdet är \bar{x} och tolkningen av medelvärdet som en tyngdpunkt enligt Figur 4.1.

För ett stickprov x_1, x_2, \dots, x_n är \bar{x} en beteckning för medelvärdet och

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

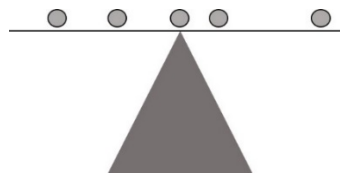
- n anger antalet mätvärden,
- stickprovets n värden betecknas x_1, x_2, \dots, x_n ,
- \bar{x} ("x-streck") är den vedertagna beteckningen för medelvärdet av ett stickprov.

Exempel 4.1.

Angreppet av virusgulst i betor i olika försöksfält anges av tabellen. Beräkna stickprovets medelvärde.

Försök nr	1	2	3	4	5
Beteckning	x_1	x_2	x_3	x_4	x_5
Virusangripen fältyta %	3,2	5,0	1,4	9,2	6,2

Medelvärdet anger "tyngdpunkten" i mätserien. Det kan illustreras med en tallinje där de enskilda mätvärdena markeras och kan t ex betraktas som vikter. Om nu en stödjepunkt placeras vid medelvärdet kommer tallinjen att fungera som en gungbräda som befinner sig i balans som i Figur 4.1.



Figur 4.1.

Tallinjen balanserar som en gungbräda med medelvärdet som stödjepunkt.

¹⁴ Det finns också andra typer av medelvärden, t ex geometriskt medelvärde och harmoniskt medelvärde, men eftersom dessa inte används här kallas genomgående det aritmetiska medelvärdet för bara medelvärdet.

Det är helt omodernt att räkna medelvärdet för hand som nedan, men det är bra för att få in beteckningarna. I praktiken lägger man in värdena i datorn och använder Excel eller ett statistiskt programpaket. I fortsättningen anges medelvärden utan beräkning för att markera att man i praktiken använder datorn för beräkningen.

Det är 5 observationer i stickprovet varför $n = 5$.

Summerar man alla värdena så får man att

$$x_1 + x_2 + x_3 + x_4 + x_5 = 3,2 + 5,0 + 1,4 + 9,2 + 6,2 = 25,0$$

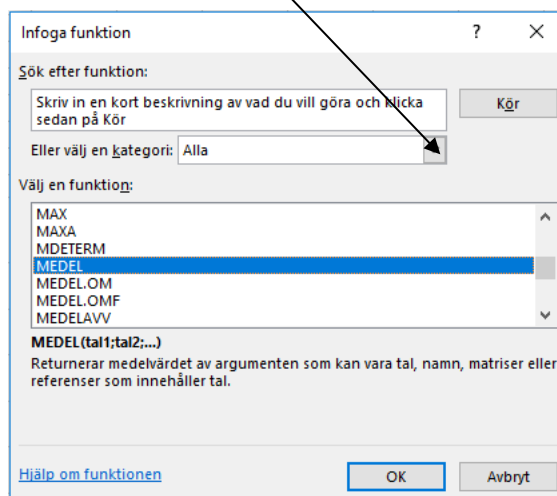
och sedan blir medelvärdet $\bar{x} = \frac{25,0}{5} = 5,0$, dvs. i medeltal är 5,0 % av ytan virusangripen.

För att göra samma beräkning i Excel läser du in värdena i en kolumn och väljer funktionen så här:

- Ställ markören i en cell där du vill ha medelvärdet. Detta skall vara en tom cell så att inte några tidigare siffror försvinner, och i detta fall är det bästa att skriva medelvärdet i den tomma cellen omedelbart under värdena på virusangripen yta (dvs. i detta fall cell B7).
- För att beräkna medelvärdet i Excel är det lättast att använda den funktionsguide som finns på formelraden och är en liten ruta märkt f_x . Funktionsguiden är mycket pedagogisk, tryck på f_x och följ de stegvisa instruktionerna

	A	B	C	D
1	Försök	Angrepp		
2	1	3,2		
3	2	5		
4	3	1,4		
5	4	9,2		
6	5	6,2		
7				
8				

- Det kommer upp en förteckning över de olika funktioner som finns. Det finns en hel del att välja på, men om du vill ha en lista på *alla* de funktioner som finns så väljer man kategorin Alla.



- Medelvärde heter i den svenska versionen MEDEL och det går att komma fortare fram i listan genom att ställa sig i listan och trycka "M". Välj OK.
- *Lägg märke till att Excel själv föreslår vilka celler som skall användas och att detta ibland inte är de celler som du vill ha!*
- För att välja vilka celler som ska användas är det lättast att markera dessa i kalkylbladet genom att trycka ned vänster musknapp och dra. För att markera placeras muspekaren *mitt i* cell B2, vänster musknapp trycks ner och dras neråt till B6 så att alla värdena är markerade. Det blir en "svag perforering" kring de celler som väljs ut.
- Tryck OK så hamnar medelvärdet i den tidigare tomma cellen. I cellen (här cell B7) står nu medelvärdet men på formelraden står formeln =MEDEL(B2:B6). Med de beteckningar som vi använt här har du nu beräknat medelvärdet \bar{x} .

	A	B	C	D	E
1	Försök	Angrepp			
2		1	3,2		
3		2	5		
4		3	1,4		
5		4	9,2		
6		5	6,2		
7			5		
8					

- Funktionen MEDEL räknar bara de celler som har värden. Det går alltså att markera tomma celler utan att det påverkar beräkningen.

Ett sammanhängande område av celler i Excel markeras med kolon (:), som anger cellerna från den första längst upp till vänster till den nedersta längst ner till höger. Det område du anger kan alltså vara ett rektangulärt område över flera kolumner t ex A2:B3.

Om du vet vad funktionen heter kan det därför ibland gå snabbare att inte använda funktionsguiden utan istället direkt på funktionsraden skriva =MEDEL(B2:B6). Glöm då inte att börja med att skriva likhetstecknet! ■

I en oändlig population blir det svårt att beräkna medelvärdet, men det finns där i teorin. För att skilja medelvärdena åt används ibland beteckningarna *stickprovsmedelvärde* och *populationsmedelvärde*.

Populationsmedelvärdet betecknas oftast med μ (grekiska bokstaven "my") och återkommer i Kapitel 5. Det är denna parameter i populationen som stickprovsmedelvärdet \bar{x} försöker uppskatta med hjälp av det slumpmässiga stickprovet.

4.3 Median

Ett annat lägesmått är *medianen* (eng. median), som är det storleksmässigt mellersta talet i mätserien. I Excel heter funktionen MEDIAN.

För att bestämma medianen *behöver talen först ordnas i storleksordning*. Medianen utgörs sedan av

- det mellersta värdet för en talserie med udda antal värden,
- medelvärdet av de två mellersta värdena för en talserie med jämnt antal värden.

Medianen har ibland vissa fördelar gentemot medelvärdet men trots det så används det ganska sällan eftersom de statistiska formlerna är uppbyggda på medelvärdet.

Exempel 4.1 (forts).

Beräkning av medianen av virusangripen fältyta i %. Först ordnas värdena i storlek från minsta till det största

1,4 3,2 5,0 6,2 9,2

och eftersom det mellersta värdet är 5,0 så är medianen 5,0 % virusangripen fältyta.

Här blir medelvärdet och medianen lika och därför kan Figur 4.1 i just detta fall också illustrera medianen som är punkten som är i mitten på gungbrädan.

Beräkningen i Excel går till på precis på samma sätt som för medelvärdet med den enda skillnaden att MEDEL i beskrivningen byts mot MEDIAN.

B7		=MEDIAN(B2:B6)		
A	B	C	D	E
1	Försök	Angrepp		
2	1	3,2		
3	2	5		
4	3	1,4		
5	4	9,2		
6	5	6,2		
7		5		
8				

Ett exempel på tillämpning av medianen är LD₅₀-värdet¹⁵, som används för att ange den akuta giftigheten hos bland annat kemiska bekämpningsmedel. Med dosen LD₅₀ borde 50 % av plantorna dö och 50 % överleva.

Ett annat exempel där medianen är att föredra är som mått på lönenivån i ett företag. Med medianen blir det en mer rättvisande bild eftersom medelvärdet kan påverkas av enstaka mycket höga löner till ett fåtal personer i företaget.

¹⁵ LD är en förkortning för Lethal Dose.

4.4 Standardavvikelse

För att beskriva datamaterialet behövs det också en parameter som anger spridningen eller variationen. Det vanligaste spridningsmåttet är *standardavvikelsen* (eng. standard deviation), som beskriver spridningen runt medelvärdet.

Även om formeln är komplicerad kan standardavvikelsen ungefär beskrivas som "medelavvikelsen från medelvärdet" och i praktiken använder man Excel eller ett statistikprogram för att räkna ut standardavvikelsen.

I Excel heter stickprovsstandardavvikelsen *STDAV.S*¹⁶.

Standardavvikelsen av ett stickprov betecknas s (ibland *SD*) och beräknas med den komplicerade formeln

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

- n anger antalet mätvärden,
- stickprovets n värden betecknas x_1, x_2, \dots, x_n ,
- \bar{x} är stickprovets medelvärde,
- nämnaren $n - 1$ kallas frihetsgrader (eng. *DF* = Degrees of Freedom) och är ett viktigt begrepp i statistiken¹⁷.

Exempel 4.1 (forts).

Beräkna stickprovsstandardavvikelsen av virusangripen fältyta.

Försök nr	1	2	3	4	5
Beteckning	x_1	x_2	x_3	x_4	x_5
Virusangripen fältyta i %	3,2	5,0	1,4	9,2	6,2

Det är absolut att rekommendera att använda datorn för att beräkna standardavvikelsen men här gör vi det för hand för att illustrera vad standardavvikelsen egentligen mäter.

Det är 5 observationer i stickprovet och därför är $n = 5$. Stickprovsmedelvärdet $\bar{x} = 5,0$ och kvadratavvikelserna som finns i formeln framgår av tabellen.

Försök nr	1	2	3	4	5
Beteckning	$(x_1 - \bar{x})^2$	$(x_2 - \bar{x})^2$	$(x_3 - \bar{x})^2$	$(x_4 - \bar{x})^2$	$(x_5 - \bar{x})^2$
Kvadratavvikelse	$(3,2 - 5,0)^2$ = 3,24	$(5,0 - 5,0)^2$ = 0,00	$(1,4 - 5,0)^2$ = 12,96	$(9,2 - 5,0)^2$ = 17,64	$(6,2 - 5,0)^2$ = 1,44

¹⁶ Det också finns en funktion som heter *STDAV.P*, och den skiljer sig från *STDAV.S* genom att den använder n i nämnaren istället för $n - 1$. Denna skall användas om man har samlat in hela populationen.

¹⁷ Det skulle här föra alltför långt att förklara varför man dividerar med $n - 1$ och inte med n när man beräknar s .

Summan av kvadratavvikelserna blir

$$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_5 - \bar{x})^2 = 3,24 + 0 + 12,96 + 17,64 + 1,44 = 35,28$$

och standardavvikelsen beräknas sedan till

$$s = \sqrt{\frac{35,28}{5 - 1}} = \sqrt{8,82} \approx 2,97$$

Standardavvikelsen har samma enhet som observationerna, i detta fall %.

Beräkningen i Excel går till på precis på samma sätt som för medelvärdet och medianen med den enda skillnaden att du använder funktionen STDAV.S.

	A	B	C	D	E
1	Försök	Angrepp			
2		1	3,2		
3		2	5		
4		3	1,4		
5		4	9,2		
6		5	6,2		
7			2,969848		
8					

Om du tycker att det är för många eller för få decimaler i resultatet i Excel så kan du ställa in detta i verktygsraden under inställningarna för tal och ikonen



Varför använda kvadratiska avvikelser? Det beror på att matematiken ofta använder kvadratisk avvikelse eftersom den funktionen är lätt att räkna med, men det är viktigt att komma ihåg att det blir väldigt känsligt för "outliers". Det syns också i handräkningen i tabellen där några av värdena i summan av kvadraterna är mycket större än de andra och påverkar standardavvikelsen mer. Ta därför för vana att rita upp ditt material för att se så att det inte finns några extrema värden.

Om populationen är oändlig så kan man inte beräkna standardavvikelsen, men den finns i teorin. Med ett slumpmässigt stickprov så kan *stickprovsstandardavvikelsen* användas för att uppskatta *populationsstandardavvikelsen*.

Populationsstandardavvikelsen är en parameter som betecknas σ (den grekiska bokstaven "sigma" motsvarar ju den svenska bokstaven s). Då det är klart av sammanhanget vilket man avser så kallas både populationsstandardavvikelsen och stickprovsstandardavvikelsen bara för standardavvikelsen.

4.5 Standard Error (medelfel)

*Standard Error*¹⁸ (*medelfel*, eng. standard error, förkortat *SE*) är ett annat ofta använt spridningsmått. Standard Error anger hur bra stickprovsmedelvärdet är som en uppskattning av populationsmedelvärdet och beräknas med formeln

$$SE = \frac{s}{\sqrt{n}}.$$

Även i svensk litteratur är det vanligast att man använder Standard Error med förkortningen *SE* som beteckning.

Det finns inte någon färdig funktion för Standard Error i Excel. Vill du beräkna det i Excel så går det att skriva formeln för hand i formelfältet om du vet att "roten ur" heter ROT och att det går att få fram antalet observationer i stickprovet med funktionen ANTAL. I det här fallet blir ju funktionen ganska komplicerad i Excel, men det fungerar. Med siffrorna från Exempel 4.1 blir resultatet

	A	B	C	D	E	F	G	H
1	Försök	Angrepp						
2	1	3,2						
3	2	5						
4	3	1,4						
5	4	9,2						
6	5	6,2						
7		1,328157						
8								

Glöm inte att börja formeln med likhetstecken så att Excel förstår att du vill göra en beräkning!

4.6 Varians

Variansen används ibland som ett alternativ till standardavvikelsen, t ex i utskrifter från Excel. Variansen definieras som standardavvikelsen i kvadrat och därför blir beteckningen för stickprovsvariansen s^2 och för populationsvariansen σ^2 .

Variansen kan beräknas med funktionen VARIANS.S i Excel.

Variansen har den stora nackdelen att den inte är i samma dimension som mätningarna. Om du mäter i *kg* så kommer variansen att ha enheten kg^2 och det gäller inte standardavvikelsen som har enheten *kg* om mätningarna är i *kg*.

¹⁸ Standard Error definieras ibland annorlunda och för att skilja definitionerna åt så kallar man ibland formeln i detta avsnitt för "Standard Error of the mean". Standard Error förväxlas ofta med standardavvikelsen.

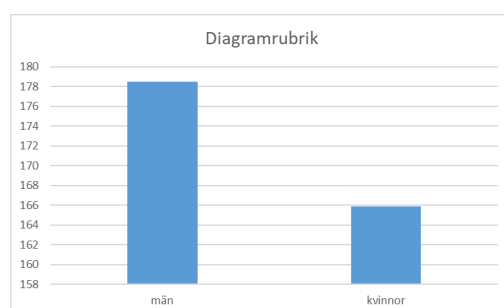
4.7 Stapeldiagram

För att grafiskt illustrera medelvärde och standardavvikelse används oftast stapeldiagram med felstaplar. För att illustrera hur detta går till använder vi ett data-material med längder hos män och kvinnor i en studentgrupp.

- Det lättaste är att flytta över siffrorna till ett tomt blad och ange vilket som är medelvärde och vilket som är standardavvikelse.

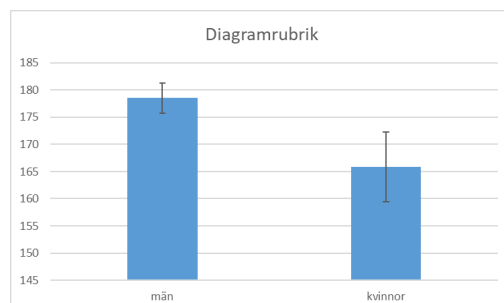
	A	B	C
1		medel	stdav
2	män	178,5	2,8
3	kvinnor	165,9	6,4
4			

- Markera nu rutorna A2:B3 (markera *inte* standardavvikelserna *men däremot texten med män och kvinnor*¹⁹) och gör ett stapeldiagram genom att välja INFOGA och bland diagrammen välja ett stapeldiagram. Detta bör ge en figur som liknar denna:



För att också illustrera standardavvikelserna lägger du in dessa som felstaplar. Det är inte helt lätt att göra detta, men här är ett försök till beskrivning:

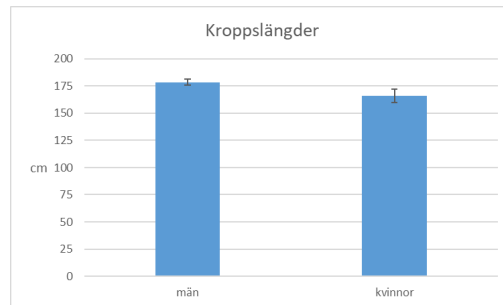
- Klicka på figuren så att den är aktiv och välj Lägg till diamelement som är en ikon som finns längst upp till vänster om du har valt menyn Design under Diagramverktyg.
- Välj Felstaplar och sedan nedersta raden med Fler Felstapelalternativ....
- Till höger kommer en meny där du väljer Anpassat, Ange värde och markera sedan standardavvikelserna för män och kvinnor (båda två samtidigt) *både i rutan positivt felvärde och i rutan negativt felvärde*. Resultatet borde nu vara



¹⁹ Detta är metoden för att få text under staplarna, det går inte att klicka på bilden i efterhand för att få dit eller redigera texten.

I nästa steg går det att bättra utseende på diagrammet genom att lägga till rubriker etc. **Du bör definitivt ändra skalan på y-axeln så att den börjar på 0 och inte är avhuggen.** Klicka därför på siffrorna på axeln för att få upp "Formatera axel" till höger. Om du har klickat på siffrorna skall det finnas en flik som ser ut som ett stapeldiagram. Klicka på den för att ändra minvärde och maxvärde i den meny som kommer upp. Vill du ändra titel och axeltext går det att göra med Lägg till diagram-element eller genom att klicka på det du vill ändra och skriva in en ny rubrik.

Här är ett förslag på en lite snyggare figur utan vilseledande skalor på y-axeln.



Männen är i medeltal längre än kvinnorna men felstaplarna visar att kvinnorna har en lite större variation eftersom standardavvikelsen är större.

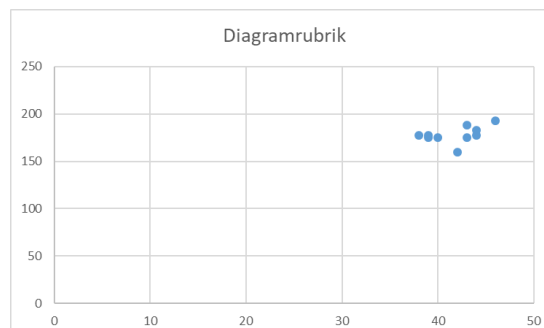
Det är viktigt att figurtexten talar om vilken spridning det är som finns i felstaplarna. Det förekommer både standardavvikelse och Standard Error som felstaplar och läsaren måste därför få reda på vilket av spridningsmått det är.

4.8 Punktdiagram

Punktdiagram används om man vill illustrera sambandet mellan två variabler, t ex längd i centimeter mot skonummer för några personer. Här använder vi skonummer och längd för tio personer.

Skonummer	38	42	39	39	46	44	43	40	44	43
Längd	177	160	175	177	193	177	175	175	183	188

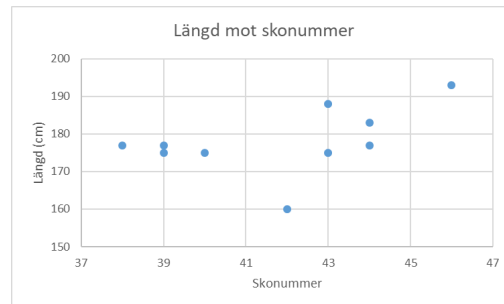
Lägg in materialet i kolumner i Excel²⁰, markera siffrorna och välj INFOGA och Punktdiagram bland de diagram som finns. Nu väljer Excel förmodligen konstiga skalor på axlarna och du får något som ser ut så här:



²⁰ Det fungerar också med rader som i tabellen.

Den lite konstiga figuren beror på att Excel i denna figur valde att börja båda axlarna på 0 och det är inte en bra idé för detta datamaterial.

Klicka därför *på siffrorna* på en axel för att få upp "Formatera axel" till höger. Om du har klickat på siffrorna skall det finnas en flik som ser ut som ett stapeldiagram. Klicka på den för att ändra minvärde och maxvärde i den meny som kommer upp. Gör sedan samma sak för den andra axeln. Redigera sedan genom att ändra Diagramrubriken och lägg till Axelrubriker för att till slut ha något som liknar den här figuren.



Det var tio personer i det här datamaterialet och det var inte någon som hade exakt samma värde. Det är viktigt att komma ihåg att det inte syns i bilden om det är flera observationer i samma punkt, om du vill illustrera det går det att använda Bubbeldiagram som vi inte behandlar i det här kompendiet.

4.9 Lådagram

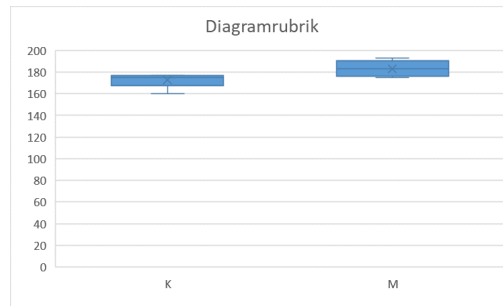
För att jämföra olika grupper av mätningar med ganska många observationer så är lådagram en bra illustration. Det engelska ordet är boxplot och det används också ofta på svenska.

I materialet med skonummer och längder i Avsnitt 4.8 fanns också en kolumn med kön och här vill vi se om längderna skiljer sig åt mellan könen.

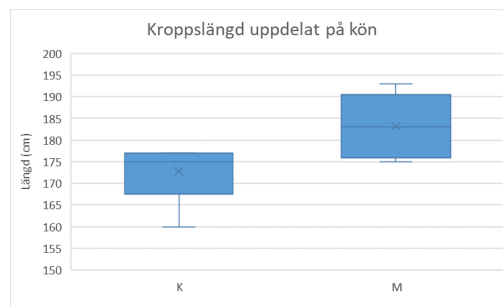
Kön	K	K	K	K	M	M	M	K	M	M
Längd	177	160	175	177	193	177	175	175	183	188

Det är inte idealt att göra lådagram med så här få observationer men fördelen är att det är bra att kunna se hela datamaterialet.

Markera datamaterialet och välj INFOGA och under en ikon som ser ut som ett stapeldiagram men heter Infoga statistiskt diagram går det att hitta Lådidiagram (det heter *ibland* så i Excel, i vissa menyer heter det Lådagram). Resultatet blir inte så bra här heller eftersom skalan börjar på 0.



Efter att ha ändrat skalan och lagt till rubriker är det lättare att se hur lådorna ser ut.

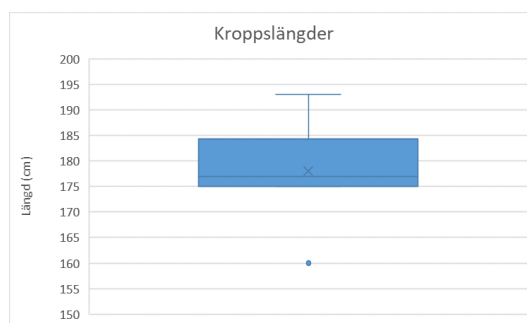


Idén med lådagrammet är att för varje grupp dela upp materialet i fyra lika stora delar; en del som är de 25 % lägsta observationerna och detta är strecken från minimum upp till lådan, en grupp med 25 % som är den undre delen av lådan och slutar i medianen, en del som är 25 % och den övre delen av lådan och den översta delen som är den övre linjen och är de 25 % högsta observationerna.

Lådan delar alltså materialet i fyra delar och den övre kanten av lådan kallas därför för den övre kvartilen, strecket i mitten är medianen och den undre kanten av lådan kallas för den undre kvartilen.

Eftersom det bara är 5 observationer i varje grupp är det svårt att dela i fyra delar, men det går att se att en kvinna är ganska mycket mindre än de andra och att medelvärdet som betecknas med \times inte helt sammanfaller med strecket för medianen.

Lådagram är bra för att skilja grupper åt, men det är också bra för att hitta "outliers" i materialet. Om hela materialet slås ihop till en grupp så kommer kvinnan som är 160 cm att anses så extrem att hon blir en punkt utanför figuren. Eftersom materialet är så skevt och litet finns det inte ett streck under lådan. Detta visar på att det kan bli lite konstigt om det är för få observationer när du gör lådagram.



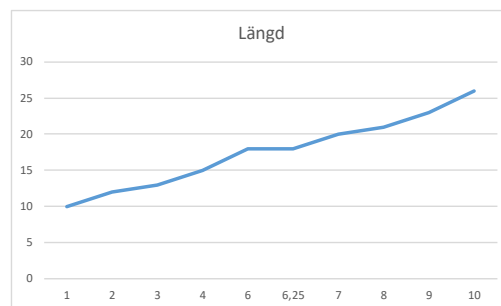
4.10 Linjediagram

Detta diagram kan vara användbart om du vill visa utvecklingen över tid, men *se upp med skalan på tidsaxeln*. Det krävs att det är samma tidsavstånd mellan alla mätningarna och att du inte saknar några observationer.

Tid	1	2	3	4	6	6,25	7	8	9	10
Längd	10	12	13	15	18	18	20	21	23	26

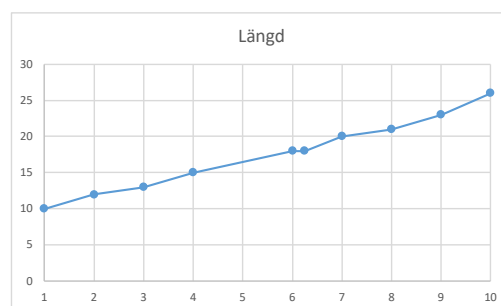
Datamaterialet är här ett tänkt material över tillväxt för en och samma planta under tio dagar. Det var tänkt att mäta vid lunch varje dag men den femte dagen fungerade inte mätapparaturen och därför blev mätningen vid lunch den sjätte dagen men också klockan 18 den sjätte dagen. Tiden för denna mätning markerades därför som tid 6,25.

Detta material ser ut att passa bra för ett linjediagram. Du får linjediagram genom att markera längderna och välja INFOGA och Linje- eller ytdiagram. Efter att ha sagt att tidsaxeln skall ha värdena från tid så blir resultatet



Linjediagrammet har alltså inte alls tagit hänsyn till att det är olika tidsavstånd mellan mätningarna.

För att inte få fel på skalan är det därför bättre att använda ett punktdiagram där du förenar punkterna med linjer. Det är bra att markera mättidpunkterna med punkter men trots att det kallas punktdiagram så går det faktiskt att ta bort punkterna och bara ha en linje.



Sammanfattningen av detta avsnitt blir att punktdiagrammet ger en ärligare beskrivning av hur plantan har utvecklats över tiden om det är olika avstånd mellan mättidpunkterna.

4.11 Histogram

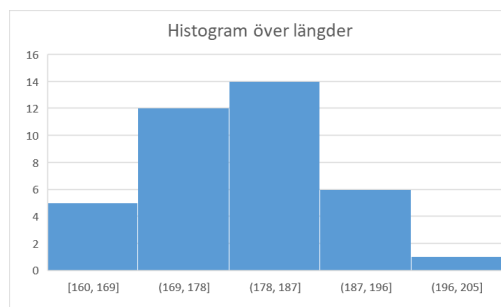
För att göra ett histogram som är meningsfullt så behövs det ganska många observationer och därför är det inte så vanligt med histogram i examensarbeten. Trots

detta beskriver vi hur du gör ett enkelt histogram utan redigeringar och hur det tolkas.

Här använder vi längderna från en grupp med 38 studenter men tabellen med värden utelämnas här.

Det finns två alternativa sätt att göra histogram, det första alternativet är att använda INFOGA som tidigare med Infoga statistiskt diagram. Här går det att välja histogram och snabbt få en figur, som dock är lite svår att redigera²¹. Det andra alternativet är att du använder histogrammet i Dataanalys som du får genom att installera Analysis ToolPak, men vi sparar det tillägget till Avsnitt 4.13 och använder därför INFOGA och Infoga statistiskt diagram

Markera materialet, välj INFOGA och hitta histogrammet bland ikonerna under Infoga statistiskt diagram. Resultatet blir



Tolkningen av histogrammet är att 5 personer har en längd mellan 160 och 169 cm. I materialet finns en person som är 178 cm²² och detta är precis i klassgränsen. Det syns inte ur histogrammet om denna person hamnar i den övre eller undre klassen. Det kan då vara bra att veta att Excel väljer att lägga observationer som hamnar i klassgränsen i den *undre* klassen.

De finns alltså mycket begränsade möjligheter att redigera detta histogram genom att ändra klassgränser, ändra skalan till relativ frekvens, etc., och det är lite lättare om man använder histogrammet i Dataanalys. Eftersom det är ovanligt att examensarbetet innehåller histogram så har vi utelämnat anvisningar om detta i det här kompendiet.

4.12 Cirkeldiagram

I vetenskapliga sammanhang så rekommenderas *inte* cirkeldiagram eftersom de lätt kan leda läsaren fel med olika färger och olika sätt att framhäva de olika tårtbitarna.

Cirkeldiagram förekommer också oftast när det är kvalitativa data och vi har därför valt att inte beskriva hur man gör cirkeldiagram i det här kompendiet.

För den som vill göra ett cirkeldiagram bör det inte vara något problem att infoga på samma sätt som vi har gjort med de andra diagrammen i det här kapitlet.

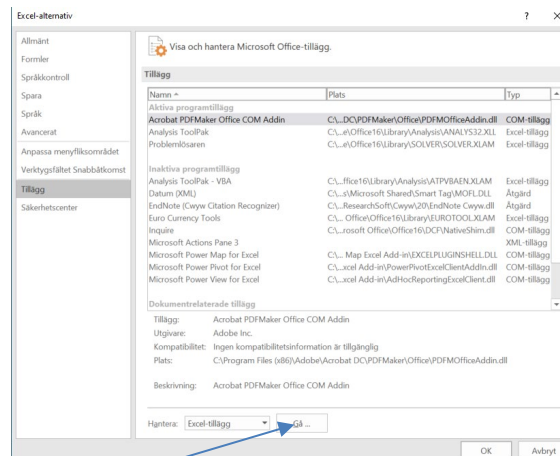
²¹ Detta kan också bero på att författaren till den här beskrivningen inte riktigt har förstått hur man redigerar det här diagrammet på ett bra sätt!

²² Tro på det, materialet finns ju inte här så att du kan kontrollera det.

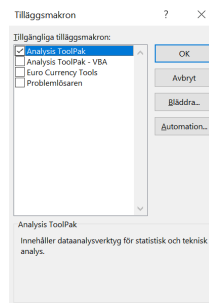
4.13 Dataanalys i Excel

För att göra lite mer avancerade statistiska beräkningar är det bra att lägga in ikonen Dataanalys. Av någon anledning finns den inte med i standarduppsättningen av Excel och du måste därför välja den bland de tillägg som finns. Här finns därför anvisning om detta²³ innan vi går vidare till Kapitel 5 där Dataanalys kommer att användas ganska mycket.

Välj menyn Arkiv, sedan Alternativ och därefter Tillägg. Du borde nu ha fått upp en sida som ser ungefär så här:

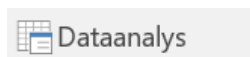


Det är nu väldigt lockande att klicka på Analysis ToolPak i listan och tro att det laddar ner Analysis ToolPak, men det skall du *inte* göra. Du skall istället klicka på rutan Gå... och få upp den här menyn



där du väljer Analysis ToolPak och sedan OK.

Om nu allt fungerar²⁴ så skall du kunna hitta en ikon som heter Dataanalys längst till höger under menyn Data.



Nu har du detta tillägg tillgängligt och som ett exempel på hur det fungerar så låter vi datorn räkna ut de flesta av de numeriska måtten i Exempel 4.1 med en enda meny.

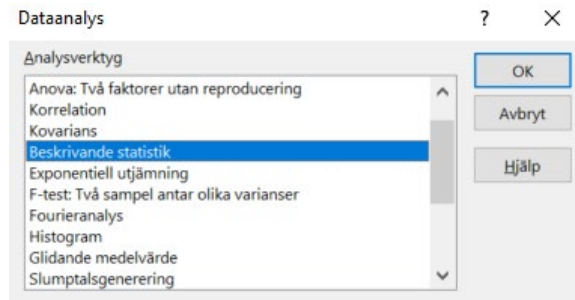
²³ Anvisningen om detta liksom andra anvisningar till Excel är för PC, men detta finns också för Mac. Leta på nätet för att se hur du skall göra om du har en Mac.

²⁴ Det har ibland hänt att det inte fungerar och vi kan inte riktigt förstå varför!

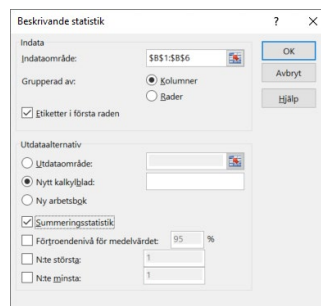
Exempel 4.1 (forts.)

Denna typ av beskrivningar av stickprovet eller populationen brukar kallas beskrivande statistik eller deskriptiv statistik (eng. descriptive statistics) och därför skall vi använda den menyn.

Klicka på Dataanalys och välj sedan Beskrivande statistik i den meny som kommer upp.



Fyll i den meny som kommer upp och om första raden innehåller variabelnamnet så kryssar du också rutan Etiketter i första raden. För att få deskriptiv statistik så kryssar du också i Summeringsstatistik.



Nu kommer du att få resultatet i ett nytt kalkylblad. I detta kalkylblad är det nästan alltid fel inställning på kolumnbredderna, men det är ju ganska lätt att ändra på.

Angrepp	
Medelvärde	5
Standardfel	1,328157
Medianvärde	5
Typvärde	#SAKNAS!
Standardavvikelse	2,969848
Varians	8,82
Toppighet	-0,09913
Snedhet	0,371076
Variationsvidd	7,8
Minimum	1,4
Maximum	9,2
Summa	25
Antal	5

Här finns nu de flesta måtten, Standardfel är detsamma som Standard Error. Typvärde är det vanligaste värdet och eftersom det var fem olika värden så saknas det värdet. Toppighet och Snedhet får betraktas som överkurs, och Variationsvidd är det som ibland kallas Variationsbredd (eng. Range) och är avståndet mellan Maximum och Minimum i datamaterialet.

Lägg märke till att kalkylbladet med resultatet *inte* är länkat till observationerna. Om du går tillbaka och ändrar någon av observationerna så kommer det *inte* att ändra sig i resultatarket, tvärt emot hur det brukar vara i Excel. ■

5 Vetenskapligt förankrade slutsatser

5.1 Definition av p-värde

I de flesta vetenskapliga artiklar som har kvantitativa data så anges p-värdet (eng. p-value) för testerna som är gjorda. Det är därför viktigt att du vet hur p-värdet tolkas, och det gäller både i dina egna studier och när du läser artiklar som andra har skrivit.

Det som gör statistiken svår är att du aldrig kan vara helt säker på att nollhypotesen är falsk, medan vetenskapsfilosofin pratar om att falsifiera. Dessa två exempel visar på en viktig skillnad mellan något som går att falsifiera och något som du aldrig helt kan falsifiera.

Slår du upp falsifiera på Wikipedia så hittar du detta exempel på något som går att falsifiera.

"Alla svanar är svarta"

Så snart du ser en vit svan så vet du att påståendet inte är sant. Om du däremot i hela ditt liv bor vid en sjö med svarta svanar så kommer du aldrig att kunna falsifiera påståendet.

Jämför dessa svanar med den statistiska nollhypotes som vi introducerade i Avsnitt 1.5.

"Behandlat och obehandlat ger samma skörd."

Vi tänker oss nu att vi samlar in 1 000 skördar av behandlat och 1 000 skördar av obehandlat och att stickprovsmedelvärdet av behandlat är mycket större än stickprovsmedelvärdet för obehandlat. Du kan trots detta aldrig vara helt säker på att behandlat skulle vara bättre än obehandlat eftersom populationerna är oändliga, du kan ju ha haft otur och fått stickprov som ger en felaktig bild av populationerna.

Detta löser statistikern genom att prata om sannolikheter, och säger ungefär så här:

Om behandlat och obehandlat skulle ha samma populationsmedelvärde så är *sannolikheten* för att få just det här resultatet så liten att du kan säga att behandlat måste ge en bättre skörd än obehandlat. Men helt säker går det aldrig att vara!

Detta är ett lite svagare krav än i juridiken där uttrycket "utom rimligt tvivel" beskriver hur säker rätten skall vara för fälla den åtalade.

Här finns alltså tre olika nivåer:

- I fallet med svarta svanar så är vi helt säkra på att det är falskt när vi har sett en vit svan.
- I en rättegång är den åtalade oskyldig tills det är ställt "utom rimligt tvivel" att detta inte kan vara sant.
- I statistiken använder man en gräns för sannolikheten (oftast 5 %) och säger att om sannolikheten att få den observerade skillnaden är mindre än den valda gränsen *om nollhypotesen är sann* så säger man att nollhypotesen förkastas.

Det är här som p-värdet löser problemet och det framgår av definitionen:

Definition

p-värdet anger hur stor sannolikheten är att få det resultat vi har fått om nollhypotesen är sann.

I så kallad "klassisk hypotesprövning" väljer du först en signifikansnivå, oftast 5 %, och ser sedan vad testet har gett för p-värde. Din slutsats blir att

- ett p-värde under eller lika med signifikansnivån 5 % innebär att du förkastar nollhypotesen till förmån för mothypotesen,
- ett p-värde över signifikansnivån 5 % innebär att du inte förkastar nollhypotesen och kan uttrycka det som att du inte har samlat tillräckliga bevis för att förkasta nollhypotesen.

Eftersom datorn så lätt skriver ut p-värdena så är det inte alltid så här det går till i praktiken. Systemet med stjärnor (*) är ett ofta använt sätt att beskriva hur mycket det går att lita på nollhypotesen.

n.s. innebär²⁵ att p-värdet är större än 0.05.

***** innebär att p-värdet är mindre än eller lika med 0.05 och större än 0.01.

****** innebär att p-värdet är mindre än eller lika med 0.01 och större än 0.001.

******* innebär att p-värdet är mindre än eller lika med 0.001.

Ibland förekommer också skrivsättet $P \leq 0,05$ för enstjärnig signifikans, $P \leq 0,01$ för tvåstjärnig signifikans och $P \leq 0,001$ för trestjärnig signifikans.

Som en avslutning på det här avsnittet måste vi tillägga att p-värdet ganska ofta kritiserats trots att det i många tidskrifter är ett krav att ha med p-värden för att artikeln skall publiceras.

Den vanligaste kritiken mot p-värdet brukar vara att

- p-värdet säger bara att det är en skillnad, men inte hur stor skillnaden är,
- om inte p-värdet är under 5 % är det väldigt svårt att få artikeln publicerad i en tidskrift,
- det är väldigt lätt att få ut ett p-värde även för den som inte kan statistik och därför finns en stor risk att p-värdet missbrukas.

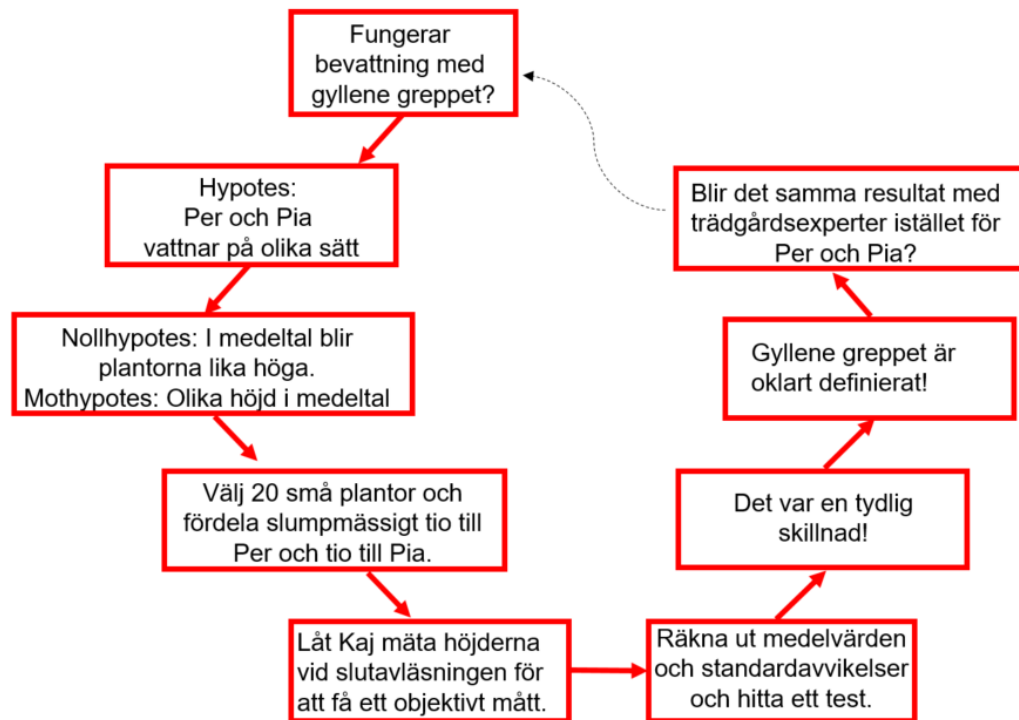
Trots den sista varningen så kommer vi i fortsättningen att fokusera på p-värdet i utskriften men samtidigt poängtera vilka förutsättningar som skall vara uppfyllda för att du skall kunna lita på dina resultat.

5.2 Arbetsgång vid hypotestest

För att belysa hela proceduren från en tanke till en färdig slutsats finns här en beskrivning i ord av hur ett påhittat experiment kan se ut. Försöket följer direkt beskrivningen av hur examensarbetet går till i Figur 5.1. Det kanske är lite svårt att smälta alla stegen direkt, men det kan vara bra att ha i bagaget inför det som kommer senare i detta kapitel.

²⁵ n.s. står för "not significant" men n.s. används också på svenska.

Per och Pia säger båda två att de använder "gyllene greppet" när de skall avgöra om plantorna skall vattnas, men Per är ganska säker på att det har betydelse om han eller Pia sköter växterna. Han vill nu försöka bevisa detta.



Figur 5.1.

Arbetsgång vid hypotestest.

- Per struntar till en början i statistiken och formulerar sin hypotes i ord och på "vanlig svenska". Hypotesen blir att Per och Pia *inte* vattnar på samma sätt trots att båda säger att de använder "gyllene greppet".
- Sedan måste han bestämma sig för hur han skall kunna se om de vattnar olika. Han tror att skillnaden är så stor att det leder till olika planthöjd vid slutavläsningen.
- Nu formulerar han de statistiska hypoteserna.
- Nollhypotesen (betecknas H_0) är i detta fall att det är samma *medelhöjd* på plantorna oberoende av om det är Per eller Pia som har vattnat, och Per hoppas kunna visa att denna nollhypotes *inte* är sann.
- Mothypotesen (betecknas H_1) beskriver vad Per tror är sant om nollhypotesen inte är sann. I detta fall är mothypotesen att det är skillnad i medelhöjd beroende på om det är Per eller Pia som vattnar.
- Per har nu formulerat om den ursprungliga hypotesen om att de vattnade olika till något som är mätbart i form av medelvärden.
- Nu måste han bestämma sig för hur stort försök han skall göra. Han vet att han måste ha upprepningar för att kunna bearbeta materialet statistiskt och

bestämmer sig för att han har tid och råd att använda 20 småplantor i ungefär samma utvecklingsstadium.

- Per ställer upp alla plantorna i rad på ett bord och väljer tio olika slumpstal mellan 1 och 20. Plantorna med dessa nummer får Pia och Per får resten.
- När han har fördelat krukorna så numrerar han sina krukor från 1 till 10 och numrerar även Pias krukor från 1 till 10.
- Per är noga med att inte något annat än vattningen skiljer plantorna åt, och därför blandar han de 20 plantorna på ett odlingsbord och flyttar ibland runt dem på bordet men är väldigt noga med att märka krukorna så att de inte förväxlar sina plantor.
- Per gör redan nu mallen till sitt försöksprotokoll i ett ark i Excel:

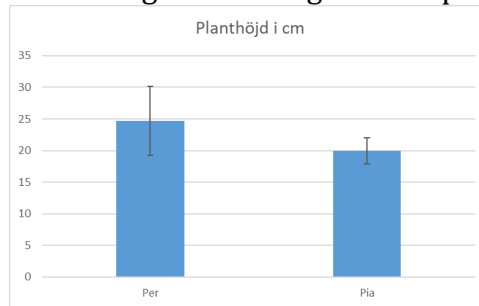
	A	B	C
1	Person	Kruka	Planthöjd
2	Per	1	
3	Per	2	
4	Per	3	
5	Per	4	
6	Per	5	
7	Per	6	
8	Per	7	
9	Per	8	
10	Per	9	
11	Per	10	
12	Petter	1	
13	Petter	2	
14	Petter	3	
15	Petter	4	
16	Petter	5	
17	Petter	6	
18	Petter	7	
19	Petter	8	
20	Petter	9	
21	Petter	10	

- Per och Pia sköter plantorna till slutavläsningen och Kaj får mäta höjderna för att få ett objektivt resultat. De bestämmer sig för att avrunda höjden till hela centimeter.
- Per för in resultaten i försöksprotokollet:

	A	B	C
1	Person	Kruka	Planthöjd
2	Per	1	22
3	Per	2	30
4	Per	3	31
5	Per	4	20
6	Per	5	25
7	Per	6	25
8	Per	7	29
9	Per	8	24
10	Per	9	28
11	Per	10	13
12	Petter	1	22
13	Petter	2	19
14	Petter	3	19
15	Petter	4	21
16	Petter	5	17
17	Petter	6	18
18	Petter	7	22
19	Petter	8	23
20	Petter	9	21
21	Petter	10	18

- Per kontrollerar snabbt så att hans observationer är rätt inlästa och att det inte är några konstiga värden. Han är lite misstänksam mot ett lågt värde (13 cm på hans egen kruka nummer 10) och kontrollerar därför denna men ser att den inte är skadad på något sätt. Han tar därför med alla plantorna i beräkningarna.
- Per räknar ut medelvärden och standardavvikelser och illustrerar resultatet i ett stapeldiagram. Det ser ut att vara en skillnad men han behöver på något sätt bevisa detta för att kunna kalla det en vetenskapligt förankrad slutsats. Per har också ganska mycket större variation på sina plantor men det

beror till viss del på att han tog med det låga värdet på sin kruka nr 10.



Nu kommer Per inte längre eftersom han inte har läst hela Kapitel 5 i det här kompendiet. Han tittar därför igenom rubrikerna och tycker att hans försöksupplägg passar med det som kallas "Två oberoende stickprov".

- I Avsnitt 5.5 hittar Per nu den statistiska metod som kallas *t*-test och använder den för att få fram ett p-värde.
- Han ställer sig i en tom cell och använder funktionen T.TEST enligt anvisningarna. Resultatet blir

=T.TEST(C2:C11;C12:C21;2;2)		
E	F	G
0,019465		

och Per vet att detta enda värde som Excel skriver ut är p-värdet för testet.

- Med ett p-värde som är 0,019 tolkar Per det som att det finns tydliga (i statistiken kallas detta *signifikanta*) skillnader mellan hans och Pias medelhöjder. Det är ju inte helt övertygande eftersom det gick att förkasta nollhypotesen på signifikansnivån 5 %, men inte på signifikansnivån 1 %.
- Hans slutsats blir att det låga p-värdet måste bero på att han och Pia har vattnat på olika sätt eftersom alla övriga faktorer hölls så lika som möjligt.
- Det fanns tillräckligt många observationer för att se skillnader mellan Pers och Pias behandlingar trots de slumpvariationer som finns med i försöket.
- Det är nu viktigt att klargöra hur mycket försöksresultatet kan generaliseras. Eftersom alla försök är gjorda på samma typ av växt och samma typ av jord är det ju inte säkert att resultatet skulle bli detsamma med en annan typ av plantjord eller en annan typ växt.

Om någon nu mer i detalj vill veta vad resultatet av testet blev så väljer Per att istället göra testet med hjälp av Dataanalys som ger mer information och inte bara p-värdet. Detta sparar vi till de senare avsnitten i det här kapitlet.

Nu fortsätter kapitlet med en genomgång av några olika försöksupplägg och om du har ett färdigt material att analysera skall du fundera på i vilket avsnitt som ditt material passar in.

Du kommer nog att ha mer än ett stickprov, men det kan ändå vara bra att läsa Avsnitt 5.3 eftersom här finns en lite mer detaljerad genomgång av hur ett test går till.

5.3 Ett stickprov

Det enklaste testet vid handräkning är att analysera ett slumpmässigt stickprov ur *en* population, och därför brukar statistikböcker börja där trots att det nästan aldrig är intressant i praktiken. Det är betydligt vanligare att jämföra två eller flera behandlingar, men vi börjar ändå med ett stickprov för att lite förklara hur ett test går till.

Det är också så att analysen i Avsnitt 5.4 som handlar om Stickprov i par bygger på de metoder som finns i detta avsnitt om ett stickprov.

Exempel 5.1.

Ett bekämpningsmedel har i medeltal lösningstiden 52 s. För att undersöka om en ny blandning har en annan lösningstid i medeltal samlar du in åtta olika slumpvis valda prov och mäter lösningstiderna med denna nya blandning. Lösningstiderna (i sekunder) blev

49,7 50,6 51,6 52,7 48,9 53,8 50,2 52,9

Först bör du alltid rita upp observationerna för att se så det inte finns något orimligt värde. Vi gör inte det här eftersom en bild av observationerna finns i Figur 5.7 som visar att det inte finns några konstiga värden bland observationerna.

Nollhypotesen är att den nya blandningen i medeltal har samma lösningstid som den gamla och det kan uttryckas som

$$H_0 : \text{”populationsmedelvärdet är 52 s”}$$

eller mer matematiskt som

$$H_0 : \mu = 52.$$

Mothypotesen kan formuleras som att den nya blandningen i medeltal har en annan lösningstid än den gamla, dvs.

$$H_1 : \text{”populationsmedelvärdet är inte 52 s”}$$

eller med matematiska beteckningar som

$$H_1 : \mu \neq 52.$$

Stickprovsmedelvärdet \bar{x} är en uppskattning av populationsmedelvärdet μ och om nollhypotesen är sann borde därför \bar{x} vara ungefär 52 sekunder. En beräkning ger att stickprovsmedelvärdet

$$\bar{x} = 51,3$$

och $51,3 - 52,0 = -0,7$. Denna skillnad mellan vad du fick och vad du trodde att du skulle få borde ju vara 0 om nollhypotesen är sann, men stickprovsmedelvärdet \bar{x} kommer självklart inte att vara exakt likt populationsmedelvärdet.

Här hjälper oss sannolikhetsteorin att avgöra hur långt ifrån 0 skillnaden skall vara innan du förkastar nollhypotesen till förmån för mothypotesen.

Sannolikhetsteorin tittar inte bara på hur stor skillnaden är, utan också på två andra viktiga saker:

- Hur många observationer finns det i stickprovet?
- Hur stor är spridningen på observationerna?

Det finns åtta observationer i stickprovet och därför är $n = 8$.

Det finns inte några tidigare försök som talar om hur stor spridningen är och därför får spridningen uppskattas med stickprovets standardavvikelse som här beräknas till $s \approx 1.727$.

I *testvärdet*, som kallas t eftersom det senare skall jämföras med en t -fördelning, vägs nu allt samman till ett standardiserat värde

$$t = \frac{\bar{x} - 52,0}{s/\sqrt{n}} = \frac{51,3 - 52,0}{1,727/\sqrt{8}} \approx -1.146.$$

Detta testvärde väger alltså samman tre saker:

1. Avvikelsen mellan stickprovsmedelvärdet \bar{x} och det värde på populationsmedelvärdet som finns i nollhypotesen finns i täljaren, i detta fall $\bar{x} - 52,0 = 51,3 - 52,0 = -0.7$.
2. Stickprovsstandardavvikelsen s , i detta fall $s \approx 1.727$.
3. Antalet observationer n , i detta fall $n = 8$.

De två sista punkterna kombineras till det som står i nämnaren, och detta är ju det som definierades som Standard Error (SE) i Avsnitt 4.5.

Om nu nollhypotesen att populationsmedelvärdet är 52 är sann så borde ju detta värde bli 0, men slumpmässiga variationer gör att det inte blir exakt 0.

Om du tittar på uttrycket för testvärdet stämmer det med hur det borde vara:

1. Stor skillnad mellan det du fick och det du borde få ger ett stort testvärde.
2. Liten spridning i materialet ger ett stort testvärde, "liten variation är bra".
3. Många observationer ger ett stort testvärde, "många observationer är bra".

Sannolikhetsteorin hjälper till genom att tala om hur långt ifrån 0 testvärdet t skall vara för att du skall ha fått en motsägelse och inte längre tror på nollhypotesen.

Det går att hitta gränser för t -värden i tabeller men i hela detta kompendium använder vi datorn för att få ett p -värde. Detta p -värde är en sannolikhet som beskriver om t -värdet är så långt ifrån 0 att det inte längre är troligt att nollhypotesen är sann, och denna gräns brukar sättas vid 5 %.

Med hjälp av ett statistiskt programpaket²⁶ kan man få fram p -värdet 0.289 vilket visar att det inte finns någon anledning att förkasta nollhypotesen (om p -värdet är större än 0.05 så finns det ingen anledning att förkasta nollhypotesen).

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
8	51,300	1,727	0,611	(49,856; 52,744)

μ : population mean of Lösningstid

²⁶ Men just detta test finns inte i Excel!

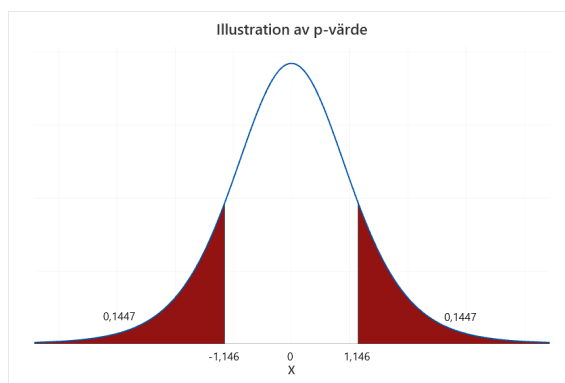
Test

Null hypothesis $H_0: \mu = 52$
 Alternative hypothesis $H_1: \mu \neq 52$

T-Value	P-Value
-1,15	0,289

Utskriften från Minitab ger all information som behövs²⁷, inklusive p-värdet. Uttrycket "SE Mean" är det som tidigare kallades Standard Error, dvs. $SE = s/\sqrt{n}$, som också är nämnaren i T .

Det beräknade p-värdet kan också illustreras med en figur. Vi använder datorn och sannolikhetsteorin för att beräkna sannolikheten att få det resultat vi fick eller något som ännu mer skulle tala för att mothypotesen är sann. Kurvan i Figur 5.2 anger hur histogrammet för t -värdet hade sett ut om nollhypotesen hade varit sann och vi upprepat försöket med åtta observationer oändligt många gånger. Sannolikheten att testvärdet hamnar i det markerade området om nollhypotesen är sann är enligt utskriften 0,289. Eftersom detta är större än 0,05 så fortsätter vi att tro på nollhypotesen.



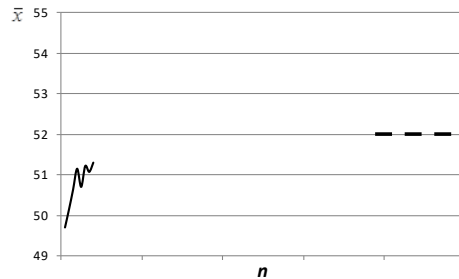
Figur 5.2.

Om nollhypotesen är sann så är sannolikheten att hamna i det skuggade området lika med p-värdet 0,289.

Ett annat sätt att formulera resultatet är att kontinuerligt uppdatera stickprovsmedelvärdet för varje ny observation och se hur det utvecklas med fler observationer. Den heldragna linjen i Figur 5.3 är stickprovsmedelvärdet av 1, 2, 3, ..., 8 observationer, dvs.

$$\begin{aligned}
 x_1 &= 49,700, \\
 \frac{x_1 + x_2}{2} &= \frac{49,7 + 50,6}{2} = 50,150, \\
 \frac{x_1 + x_2 + x_3}{3} &= \frac{49,7 + 50,6 + 51,6}{3} = 50,633, \\
 &\vdots \\
 \frac{x_1 + \dots + x_8}{8} &= \frac{49,7 + \dots + 52,9}{8} = 51,300.
 \end{aligned}$$

²⁷ I Avsnitt 5.6 beskrivs vad konfidensintervallet (CI) är för något.



Figur 5.3.

Illustration av resultatet. Det är *inte* omöjligt att kurvan kommer att svänga in mot den streckade linjen till höger eftersom p -värdet är större än 5 %.

Det är viktigt att komma ihåg att det går att förkasta nollhypotesen även om den är sann, och risken anges av signifikansnivån. Om signifikansnivån är 5 % och det inte finns någon skillnad så kommer du ändå i genomsnitt att säga att det är skillnad i vart 20:e försök. Det finns därför en hel del publicerade resultat som tvärsäkert talar om effekter när det i själva verket bara är slumpmässiga orsaker som har gjort att p -värdet blev under 5 %.

Om testvärdet, i detta fall t , inte är tillräckligt stort/litet för att förkasta nollhypotesen betraktas försöksresultatet som inte signifikant. Gränsdragningen följs väldigt konsekvent, vilket innebär att det i vetenskapliga artiklar *inte* är tillåtet att prata om skillnader eller effekter om de inte är signifikanta.

Om nu försöksresultat inte är signifikant kan nollhypotesen inte förkastas. Det innebär enkelt uttryckt att sannolikheten är större än 5 % att slumpen kan ha orsakat resultatet.

För att sammanfatta detta går det att se två orsaker till att testet inte kan förkasta nollhypotesen:

- Nollhypotesen är sann.
- Nollhypotesen är inte sann men slumpeffekten är så stor att det inte går att påvisa detta med det slumpmässiga stickprov som finns i experimentet.

Testet som gjordes i Exempel 5.1 var ett så kallat tvåsidigt test eftersom mothypotesen innehöll värden både över och under 52. Detta innebär att både stora positiva och stora negativa värden på testvärdet t medför att nollhypotesen förkastas.

I vissa sammanhang är det naturligt att använda en mothypotes som bara innehåller värden som är större än (eller mindre än) nollhypotesens värde. I Exempel 5.1 är syftet ganska säkert att göra en blandning som minskar lösningstiderna, och då är det inte intressant att visa att det är en signifikant ökning av lösningstiderna. Ett positivt värde på testvärdet t skulle därför bara innebära att du absolut inte tror att den nya metoden ger kortare lösningstider. För att visa att det är i stort sett samma beräkningar så räknas Exempel 5.1 om med en ensidig mothypotes.

Exempel 5.1 (forts).

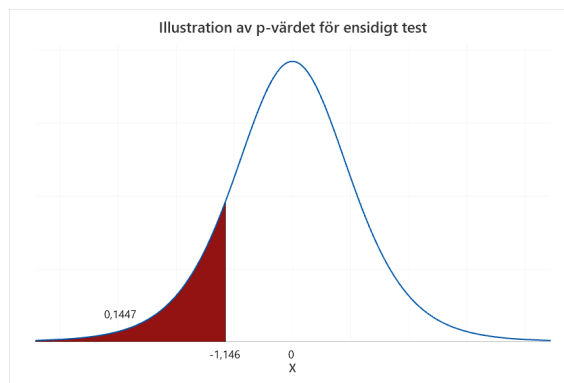
Eftersom man tror att lösningstiderna är kortare än 52 s så formuleras hypoteserna enligt²⁸

$$H_0 : \text{”populationsmedelvärdet är 52 s”}$$

mot

$$H_1 : \text{”populationsmedelvärdet är mindre än 52 s”}$$

Testvärdet beräknas exakt som tidigare och blir därför $t = -1,146$. Om nu verkligen den nya metoden ger lägre värden så borde testvärdet t vara negativt och det som talar ännu mer för mothypotesen är nu bara värden som är mindre än $-1,146$ långt ut i fördelningens svans på den negativa sidan. Om testvärdet är långt ut i svansen på den positiva sidan talar ju detta bara för att den nya blandningen ger längre lösningstider.



Figur 5.4.

Illustration av p-värdet i exemplet med ensidig mothypotes. ■

I praktiken glöms ofta att det egentligen borde vara ett ensidigt test eftersom datorprogrammen automatiskt levererar tvåsidiga test om du inte anger något annat.

Det är viktigt att bestämma sig för att göra ett ensidigt test *innan* datamaterialet är insamlat, eftersom det annars går att välja mothypotesen i den riktning som datamaterialet visar och därmed sänka p-värdet på ett otillåtet och ovetenskapligt sätt.

5.4 Stickprov i par

Exempel 5.1 har behandlat *ett* stickprov och i Avsnitt 2.4 stod också att blockförsök med bara två behandlingar testas med samma teknik. Detta brukar kallas stickprov i par eller alternativt matchat t -test. Ett exempel visar hur detta går till.

²⁸ Nollhypotesen hade också kunnat skrivas som att populationsmedelvärdet är större än eller lika med 52 s utan att det hade förändrat beräkningarna.

Lägg märke till att beräkningen fordrar att försöket är upplagt som ett blockförsök. Det innebär i detta fall att värdena hör ihop parvis, om det inte är så får du räkna enligt de metoder som anges för två stickprov i Avsnitt 5.5.

Exempel 5.2.

Försöksplan, rutfördelning och skörd från ett fältförsök med endast två led. Försöket är upplagt som ett blockförsök (varje block med en uppsättning led).

A = Obehandlat

B = Behandlat, t ex mot bladlöss vid förekomst (före en lus/strå).

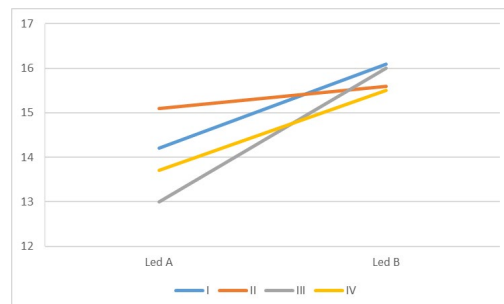
Rutfördelning: Romerska siffror anger block.

I		II		III		IV	
A	B	B	A	B	A	A	B

Försöket gav dessa rutskördar (i kg):

I		II		III		IV	
14,2	16,1	15,6	15,1	16,0	13,0	13,7	15,5

Det verkar som att behandlingen har haft en effekt, Figur 5.5 visar att det har blivit merskörd vid behandlingen i alla fyra blocken. Frågan är om skillnaden är så stor att det går att belägga det med statistiska metoder.



Figur 5.5.

Skördarna verkar vara bättre för de behandlade rutorna.

För att illustrera den fortsatta beräkningen görs en ny tabell med differenser, men det är i praktiken helt onödigt eftersom datorn beräknar differenserna. Om man räknar differensen B–A eller A–B spelar ingen roll, ofta anpassar man det så att man får ett positivt stickprovsmedelvärde för differenserna.

Block	Led A	Led B	Differens B–A
I	14,2	16,1	1,9
II	15,1	15,6	0,5
III	13,0	16,0	3,0
IV	13,7	15,5	1,8

Efter att ha beräknat dessa differenser är det dessa som gäller i den fortsatta beräkningen. Det innebär att det blir ett test för *ett* stickprov som vi beskrev i Avsnitt 5.3. Om det inte är någon skillnad mellan A och B så skall differensen i medeltal vara 0, och det får därför bli testets nollhypotes.

Eftersom vi inte vet om B förväntas ge en högre eller lägre skörd blir mothypotesen tvåsidig. Hypoteserna blir därför

H_0 : "populationsmedelvärdet för differenserna är 0"

mot

H_1 : "populationsmedelvärdet för differenserna är inte 0".

För att markera att det bara är differenserna som nu ingår i beräkningen skrivs stickprovets värden ut:

$$x_1 = 1,9, \quad x_2 = 0,5, \quad x_3 = 3,0, \quad x_4 = 1,8.$$

Ur dessa beräknas stickprovsmedelvärdet $\bar{x} = 1,8$, vilket innebär att behandlingen har resulterat i en skördeökning på i genomsnitt 1.8 kg per block.

Formeln för testvärdet t , baserat på differenserna, blir analogt med Exempel 5.1 men nu baserat på standardavvikelsen av differenserna, $s = 1,023$ och $n = 4$ eftersom det är fyra differenser.

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{1,8 - 0}{1,023/\sqrt{4}} \approx 3,52.$$

För att avgöra om 3,52 är långt ifrån 0 så går det att använda funktionsguiden i Excel på de ursprungliga observationerna för att få fram p-värdet. Det framgår av utskriften att man i menyn skall välja 2 för sidor eftersom det är tväsidigt test och 1 för typ eftersom det är ett parat test.

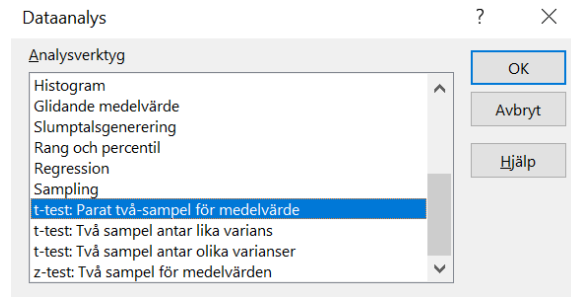
	A	B	C	D	E	F
1	Block	Led A	Led B			
2	I	14,2	16,1		0,03895	
3	II	15,1	15,6			
4	III	13	16			
5	IV	13,7	15,5			
6						

Slutsatsen av detta blir att p-värdet beräknas till 0.039, dvs. 3,9 % och att

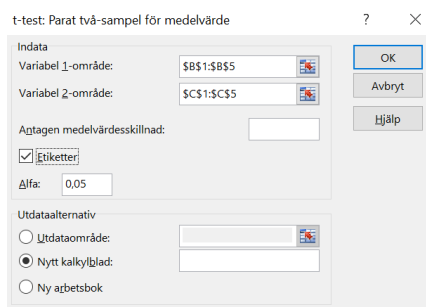
- om nollhypotesen är sann så är sannolikheten att få detta resultat eller något som ännu mer hade talat för mothypotesen mindre än 5 %,
- eftersom sannolikheten är mindre än 5 % förkastas nollhypotesen.

Det finns en signifikant skillnad mellan behandlingarna, men den är inte helt övertygande eftersom det bara gick att förkasta hypotesen på signifikansnivån 5 % och inte på lägre signifikansnivåer. Detta beror till mycket stor del på att det var så få observationer i försöket.

Om du vill använda Excel men få ut mer än ett p-värde i en cell så kan du använda Datanalys från Analysis Toolpak. Välj



och markera de celler där du har dina värden. *Om du har tagit med namnen i översta raden så skall du kryssa för rutan Etiketter.*



Resultatet kommer nu i ett eget kalkylblad om du inte anger något annat. Lägg märke till att detta *inte* är länkat till observationerna. Om du går tillbaka och ändrar någon av observationerna så kommer det *inte* att ändra sig i resultatarket, tvärt emot hur det brukar vara i Excel.

Kolumnbredderna är oftast fel i utskriften, så för att få det läsbart är kolumnbredderna här justerade så att texten får plats

t-test: Parat två-sampel för medelvärde		
	Led A	Led B
Medelvärde	14	15,8
Varians	0,78	0,086666667
Observationer	4	4
Pearson-korrelation	-0,346153846	
Antagen medelvärdesskillnad	0	
fg	3	
t-kvot	-3,518830147	
P(T<=t) ensidig	0,019474978	
t-kritisk ensidig	2,353363435	
P(T<=t) tvåsidig	0,038949955	
t-kritisk tvåsidig	3,182446305	

Tolkning av utskriften:

- Medelvärdena är självklart intressanta.
- Varianserna är standardavvikelseerna upphöjt till 2 enligt Avsnitt 4.6. För att få standardavvikelseerna för respektive led används funktionen ROT.
- Testvärdet t heter här t-kvot och avrundas till -3.52 som i handräkningen.
- p-värdet anges både för ensidig och tvåsidig mothypotes, men i detta fall var det tvåsidig mothypotes och därför blir p-värdet avrundat till 0,039.

Övrigt som står i tabellen är inte så intressant i det här sammanhanget och du ser ju också att det är onödigt många decimaler, men det går ju att justera med ikonen Minska decimal. ■

Analysen gick alltså ut på att beräkna differenser för varje par, men hur redovisar man resultatet i en figur eller i en tabell?

Det vanligaste är att använda de värden som finns i utskriften trots att felstaplarna är missvisande eftersom de också innehåller variationen mellan blocken.

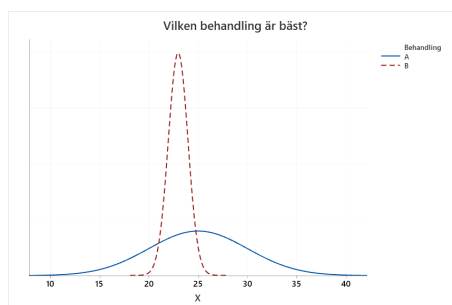
För att göra ett stapeldiagram med resultatet av försöket samlar man därför medelvärdena och standardavvikelsena intill varandra som det illustrerades i Avsnitt 4.7 och gör stapeldiagrammet enligt beskrivningen som finns där.

5.5 Två oberoende stickprov

Om det är två behandlingar men försöket inte är upplagt som ett blockförsök så går det inte att para ihop rutorna och bilda differenser, och du måste då använda en annan metod.

Det är nu två stickprov och det finns en uppskattning av standardavvikelsen från varje stickprov. Du måste då bestämma dig för om du skall anse att det är samma spridning i de båda populationerna eller inte. *Det absolut vanligaste är att anta att populationsstandardavvikelsena är lika och därför är det den enda metod som vi beskriver här.*

Om du använder en dator vid beräkningarna är det väldigt lätt att ändra detta och säga att populationsstandardavvikelsena inte är lika, men då kan man hamna i lite diskussion om vad som egentligen är den bästa behandlingen. Även om medelvärdet för en behandling är störst så kanske den inte är användbar om spridningen mellan observationerna är mycket större än i den andra behandlingen.



Figur 5.6.

Med olika standardavvikelse är det inte bara medelvärdet som avgör vilken population som är bäst.

För att ge ett exempel på hur man analyserar två oberoende stickprov används samma siffror som i Exempel 5.2, men för att markera att de kommer från en annan försöksplan finns här hela proceduren från början. Då framgår det också vilka skillnader och likheter som finns i beräkningarna.

Exempel 5.3.

Försöksplan, rutfördelning och skörd från ett fältförsök med endast två led. Försöket är upplagt som ett balanserat och fullständigt randomiserat försök.

A = Obehandlat

B = Behandlat, t ex mot bladlöss vid förekomst (före en lus/strå).

Randomiseringen valde fyra siffror mellan 1 och 8 för att bestämma vilka rutor som skulle vara behandlade. De fyra som inte valdes ut blev obehandlade. De fyra dragna siffrorna blev 6, 5, 2, 4 och rutfördelningen

A	B	A	B	B	B	A	A
---	---	---	---	---	---	---	---

Försöket resulterade sedan i följande rutskördar (i kg):

14,2	16,1	15,1	15,6	16,0	15,5	13,0	13,7
------	------	------	------	------	------	------	------

Du bör nu rita upp observationerna för att se att det inte finns några konstiga värden, men vi har redan ritat upp dem i Figur 5.5.

För att fortsätta beräkningen behövs några nya beteckningar.

Låt μ_A vara populationsmedelvärdet av skördarna från behandling A, och låt μ_B vara populationsmedelvärdet av skördarna från behandling B. Båda behandlingarna antas ha *samma spridning* på skörderesultaten, och därför finns det en gemensam populationsstandardavvikelse som brukar ha beteckningen σ .

Med dessa nya beteckningar kan nollhypotesen att det inte är någon skillnad i medelvärde mellan A och B uttryckas som

$$H_0: \mu_A = \mu_B$$

och mothypotesen att det är en skillnad mellan behandlingarna kan uttryckas som

$$H_1: \mu_A \neq \mu_B$$

Om du inte har dator beräknar du nu testvärdet t , som har en lite mer komplicerad formel än vid ett stickprov. Vi skriver därför inte ut formeln för testvärdet t här men den innehåller

1. Differensen $\bar{x}_A - \bar{x}_B$ mellan medelvärdena som borde vara 0 om nollhypotesen är sann.
2. Den uppskattade gemensamma standardavvikelsen som här måste vägas ihop med avseende på de båda stickprovens storlek och stickprovsstandardavvikelser. Detta brukar kallas den poolade standardavvikelsen som betecknas s .
3. Antalet observationer från de båda stickproven n_A respektive n_B . Ju fler observationer du har, ju lättare är det att hitta en signifikant skillnad mellan behandlingarna.

Ur detta beräknas sedan testvärdet t och precis som tidigare bör detta vara nära 0 om nollhypotesen är sann. För att se om testvärdet är extremt långt från 0 finns p-värdet som säger om nollhypotesen skall förkastas eller inte.

Excel kan hjälpa oss att få ut p-värdet, både med funktionsguiden som bara ger p-värdet och med Dataanalys som ger lite mer information, bland annat testvärdet.

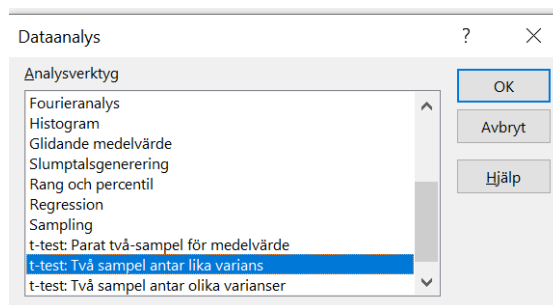
Med funktionsguiden blir det som tidigare bara ett p-värde utskrivet i cellen. Det framgår av utskriften att man i menyn skall välja 2 för sidor eftersom det är tvåsidigt test och 2 för typ eftersom det är "två sampel med lika varians (homoscedastisk)".

	A	B	C	D	E	F
1	Led A	Led B				
2	14,2	16,1		0,008294		
3	15,1	15,6				
4	13	16				
5	13,7	15,5				
6						

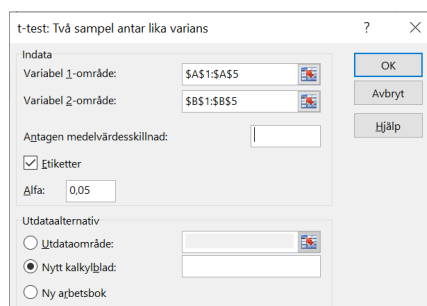
Resultatet blir alltså här att p -värdet är 0,008. Nollhypotesen kan förkastas på signifikansnivån 1 % (p -värdet är ju mindre än 0,01) men inte på signifikansnivån 0,1 % (p -värdet är ju större än 0,001). Detta skulle också kunna uttryckas som att det är ** signifikant.

- Om nollhypotesen är sann så är sannolikheten att få det resultat man fick eller något som ännu mer skulle tala för mothypotesen mindre än 1 %.
- Slutsatsen blir att det finns en signifikant skillnad mellan behandlingarna och nollhypotesen kan förkastas på signifikansnivån 1 % men inte på lägre nivåer.

Om du skall göra motsvarande beräkning i Dataanalys så väljer du analysverktyget



och markerar sedan ditt datamaterial (om variabelnamnen är med kryssar du i rutan Etiketter), men du behöver inte skriva något i rutan Antagen medelvärdesskillnad eftersom programmet antar att det är 0 om du inte anger något annat.



Efter att ha justerat kolumnbredden blir resultatet

t-test: Två sampel antar lika varians		
	Led A	Led B
Medelvärde	14	15,8
Varians	0,78	0,086666667
Observationer	4	4
Parad varians	0,433333333	
Antagen medelvärdeskillnad	0	
fg	6	
t-kvot	-3,86702032	
P(T<=t) ensidig	0,00414724	
t-kritisk ensidig	1,943180281	
P(T<=t) tvåsidig	0,008294479	
t-kritisk tvåsidig	2,446911851	

Här finns en del överflödigt information och återigen onödigt många decimaler. Om du vill göra ett stapeldiagram med medelvärden och standardavvikelser så finns detta här om du kommer ihåg att standardavvikelsen är roten ur variansen och funktionen =ROT beräknar detta. Här får man också ett testvärde t på raden för t-kvot men det mest intressanta värdet är ju p-värdet för tvåsidigt test som här är detsamma som i den tidigare metoden, dvs. 0,008294.

Det vanligaste sättet att illustrera resultatet är med stapeldiagram för medelvärdena och felstaplar som anger standardavvikelserna. I Avsnitt 4.7 hittar du en beskrivning av hur detta går till. ■

Jämför du nu Exempel 5.2 och Exempel 5.3 så verkar det ju dumt att göra ett blockförsök eftersom p-värdet blev lägre när vi använde ett fullständigt randomiserat försök. Blockindelningen var en säkerhetsåtgärd om det skulle vara en skillnad i skörd i väst-östlig riktning. Nu var denna skillnad inte så tydlig men det gick ju inte att veta innan försöket utfördes. Blockindelning görs som regel för att gardera sig mot stora variationer över försöksytan, och *ett försök som är randomiserat som ett blockförsök bör också analyseras som ett blockförsök även om det inte går att se tydliga blockeffekter.*

5.6 Konfidensintervall för medelvärden

I det här avsnittet beskrivs konfidensintervall som är ett annat sätt att undersöka medelvärden och skillnader mellan medelvärden.

Hypotestestet har nackdelen att bara tala om att det finns en eventuell skillnad och inte hur stor denna skillnad är. Konfidensintervallen som kort presenteras i detta avsnitt har inte denna nackdel och har också fördelen att du inte behöver ange några hypoteser.

Det som talar emot att använda konfidensintervall är att Excels funktioner helt är uppbyggda på hypotestest och att du får använda andra program om du vill ha konfidensintervall.

Idén är helt enkelt att göra ett intervall som täcker det rätta värdet på populationsmedelvärdet med en viss angiven sannolikhet (=konfidensgrad).

Det finns både ensidiga och tvåsidiga konfidensintervall men för att göra det lite lättare så beskrivs här bara tvåsidiga konfidensintervall, vilket förekommer mest i praktiken. Det är också bara intervall med konfidensgraden 95 % eftersom det är den klart mest använda konfidensgraden i praktiken.

För *ett* stickprov ger datorn ett intervall som täcker populationsmedelvärdet μ med 95 % sannolikhet. Detta konfidensintervall har medelpunkten i stickprovsmedelvärdet \bar{x} och intervallets bredd beror på hur stor spridning och hur många observationer det är i stickprovet.

Hur tolkas nu intervallet? Det säger att med 95 % säkerhet, i 19 fall av 20, kommer populationsmedelvärdet μ att täckas av det angivna intervallet. Om man hela tiden använder konfidensgraden 95 % så kommer i medeltal 19 intervall av 20 att täcka det rätta värdet, i medeltal 1 intervall av 40 att vara fel genom att båda gränserna ligger under det rätta värdet och i medeltal 1 intervall av 40 att vara fel genom att båda gränserna ligger över det rätta värdet.

Exempel 5.1 (forts).

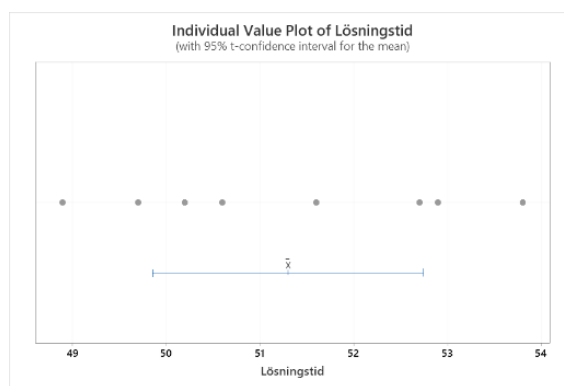
I Excel får du inte konfidensintervallet i utskriften så vi tar det istället från Minitab. Konfidensintervallet med 95 % konfidensgrad är angivet till (49,9; 52,7). Den nedre gränsen är alltså 49,9 och den övre gränsen är 52,7.

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
8	51,300	1,727	0,611	(49,856; 52,744)

μ : population mean of Lösningstid

Ber du också Minitab att göra en "Individual Value Plot" ritas konfidensintervallet ut som i Figur 5.7. Lägga märke till att fyra av observationerna är utanför konfidensintervallet. Det är en ganska vanlig missuppfattning att konfidensintervallet innehåller 95 % av populationens värden. Så är det inte alls, konfidensintervallet är ett intervall som med 95 % sannolikhet täcker populationsmedelvärdet. Med fler observationer så blir intervallet smalare och då hamnar fler och fler observationer utanför intervallet.



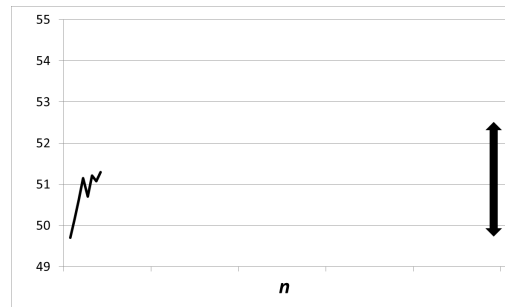
Figur 5.7.

Datamaterial och konfidensintervall med 95 % konfidensgrad från Minitab.

Det finns ett väldigt nära samband mellan konfidensintervall och hypotestest som inte blir så tydligt när formlerna inte är utskrivna i detalj: Om intervallet med 95 % konfidensgrad täcker värdet i nollhypotesen så går det inte att förkasta nollhypotesen på nivån 5 % och omvänt, om intervallet inte täcker värdet i nollhypotesen så kan nollhypotesen förkastas. Konfidensintervallet kan sägas innehålla "rimliga

gissningar” på populationsmedelvärdet och de rimliga gissningarna är de värden som *inte* förkastas i ett hypotestest.

En lite annorlunda tolkning av konfidensintervallet finns i Figur 5.8. Det är samma bild av hur stickprovsmedelvärdet ”svänger in” som fanns i Figur 5.3. Nu är frågan var det kommer att hamna i oändligheten och konfidensintervallet säger att med 95 % sannolikhet så kommer man att hamna mellan de två gränser som ges av dubbelpilen som anger intervallet.



Figur 5.8.

Illustration av konfidensintervall med konfidensgraden 95 %.

Lägg speciellt märke till att konfidensintervallet säger var **populationsmedelvärdet** ligger, det säger *inte* att 95 % av observationerna ligger inom konfidensintervallet.

För två stickprov anger konfidensintervallet vad skillnaden mellan populationsmedelvärdena (dvs. $\mu_A - \mu_B$) är med 95 % konfidensgrad.

Exempel 5.3 (forts).

Excel är helt baserat på hypotestest och om du vill ha konfidensintervall får du använda ett annat program och här använder vi Minitab.

Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
-1,800	0,658	(-2,939; -0,661)

Konfidensintervallet med 95 % konfidensgrad alltså av $(-2,94; -0,66)$.

Eftersom detta intervall *inte* innehåller 0 så har man också visat att hypotestestet som testar om behandlingarna är lika kan förkasta nollhypotesen på signifikansnivån 5 %. Konfidensintervallet innehåller rimliga gissningar på differensen och eftersom 0 inte är en rimlig gissning så är också nollhypotesen att differensen är 0 orimlig och kan förkastas. Detta stämmer ju också med att p-värdet i detta exempel var under 5 %.

Eftersom intervalllets båda gränser är negativa och differensen är $A - B$ så antyder det ju också att B har högre medelvärde.

Detta sätt att jämföra behandlingar genom att se på konfidensintervallet för skillnaden används i senare avsnitt i det här kapitlet.

Om man ser uttryck av typen " $\bar{x} \pm$ något" så skall man noga titta efter vad "något" är. Det finns nämligen tre alternativ som alla förekommer:

- Konfidensintervall beräknat enligt den metod som angetts i detta avsnitt.
- $\bar{x} \pm SE$, där SE står Standard Error (= medelfel).
- $\bar{x} \pm s$, där s är standardavvikelsen som ibland också betecknas SD .

5.7 Mer än två behandlingar

I försök med mer än två behandlingar skulle du kunna använda t -testet från tidigare avsnitt i kapitlet för alla par av försöksled, men det finns tre saker som talar emot detta:

- Om du har många behandlingar så blir det väldigt många t -tester, med sex behandlingar har du 15 par av behandlingar.
- Du utnyttjar inte datamaterialet på bästa sätt i förhållande till den modell du har.
- Om alla de parvisa testerna har signifikansnivån 5 % och nollhypotesen är att alla behandlingarna är lika så blir signifikansnivån mycket större än 5 % när du gör alla de parvisa testerna.

För att komma runt de här problemen så använder du istället *variationsanalys* (eng. analysis of variance, ofta förkortat ANOVA²⁹).

I de enklaste fallen av variationsanalys går det att beräkna de värden som behövs för hand, men här används datorn helt och hållet. Istället för handräkningar läggs tyngdpunkten på att förstå vilka förutsättningar som skall vara uppfyllda för att du skall kunna göra en variationsanalys samt att tolka de resultat som datorn producerar.

Det är viktigt att notera att ett grundläggande antagande i modellen säger att standardavvikelserna är desamma för alla behandlingarna, eftersom det inte är självklart att detta villkor är uppfyllt. Har du t ex en kontroll som ger väldigt dålig skörd så kommer förmodligen denna att ha en mindre variation än de övriga behandlingsleden. Trots detta är det väldigt ovanligt att ändra modellen och säga att det är olika spridning, lite beroende på det problem med tolkningen som har illustrerats i Figur 5.7.

Excel är väldigt begränsat när det är mer än två behandlingar, men vi visar ändå hur du kan använda Excel på de problem som finns i Avsnitt 5.8 och Avsnitt 5.9. Har du ett material med mer än två behandlingar så rekommenderar vi definitivt att du använder ett annat statistikprogram för att få fram de statistiska resultaten eftersom de resultat som finns i Avsnitt 5.12 och Avsnitt 5.13 inte går att beräkna med Excel.

I det här kapitlet så har vi använt Minitab, men det går lika bra att använda andra statistikprogram. Programmet R med RStudio som kan laddas ner gratis har blivit alltmer populärt och användarvänligt genom att det hela tiden byggs på med nya paket.

Inom växtodling och växtskydd är det definitivt vanligast med försök utlagda i block, men det är enklare att börja med ett fullständigt randomiserat försök utan

²⁹ Variansen är ett spridningsmått som beskrevs i Kapitel 4 (standardavvikelsen i kvadrat).

block och spara blockförsöket till Avsnitt 5.9. Variansanalysen som avgör om det är några skillnader mellan de olika behandlingarna kan göras i Excel, men senare avsnitt i kapitlet behandlar den ännu viktigare frågan om vilka behandlingar som skiljer sig åt.

5.8 Variansanalys utan blockindelning

Om du inte har block och bara har en faktor att studera så brukar försök av denna typ kallas *envägs ANOVA*. I Kapitel 2 om försöksplanering var rådet att du skulle sträva efter att ha balanserade försök, dvs. att alla behandlingar har lika många observationer. När man har en modell med envägs ANOVA blir beräkningarna inte så mycket mer komplicerade om man har obalanserade försök, men för att inte göra det onödigt komplicerat så använder vi här balanserade försök.

Exempel 5.4.

För att jämföra fyra behandlingar (A, B, C, D) delades 16 försöksrutor slumpmässigt upp på fyra grupper med fyra rutor i varje, en grupp för varje behandling. Skörden i de olika rutorna framgår av tabellen.

Behandling			
A	B	C	D
25	21	24	23
23	23	23	25
20	20	24	21
21	19	26	25

Nollhypotesen är här att de olika behandlingarna *i medeltal* ger samma skörd.

I en formel kan lika populationsmedelvärde uttryckas som

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D$$

eller

$$H_0: \text{populationsmedelvärdena är lika}$$

Mothypotesen är att det någonstans finns en skillnad mellan behandlingarnas medelskördar, dvs.

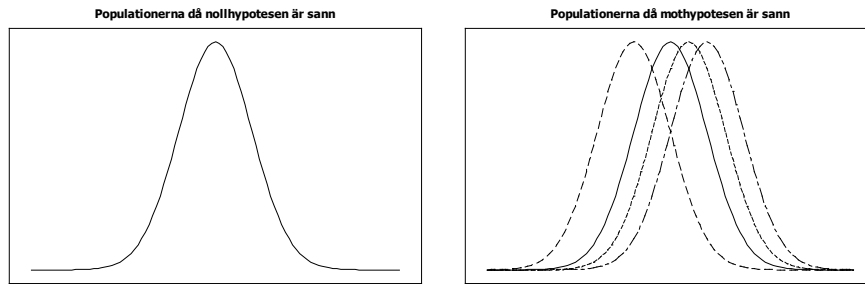
$$H_1: \text{"alla de fyra populationsmedelvärdena är inte lika"}$$

Några figurer kan illustrera vad hypoteserna handlar om. Variansanalysen antar att observationerna är normalfördelade med samma standardavvikelse för alla populationerna (eller åtminstone att detta antagande inte är helt orimligt). Om du skulle kunna samla in oändligt många försöksrutor från varje behandling och dessutom haft fler decimaler i mätningarna så hade histogrammen blivit fyra kontinuerliga kurvor som hade visat hur de fyra olika populationerna ser ut.

Om nollhypotesen att behandlingarna är lika är sann så hade allt hamnat i samma kurva enligt kurvan till vänster i Figur 5.9.

Om det däremot finns en skillnad mellan de olika medelvärdena så hade histogrammen istället sett ut som kurvorna till höger i Figur 5.9. Lägg märke till att detta bara är en variant av hur mothypotesen kan se ut, det kan ju vara så att A, B

och C har samma skörd men att D skiljer sig från de övriga. I det fallet så skulle tre kurvor sammanfalla och den fjärde ligga på ett annat ställe.

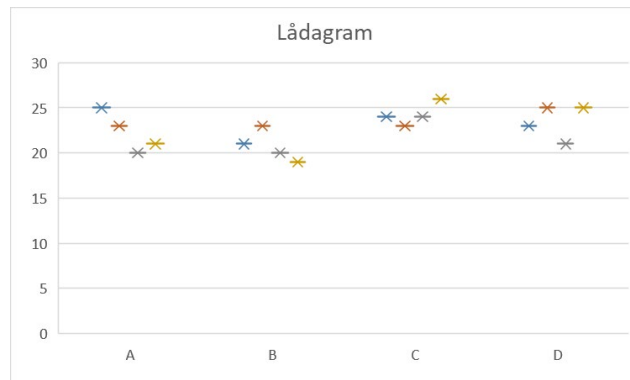


Figur 5.9.

Illustration då nollhypotesen respektive mothypotesen är sann.

Om du nu plockar ut ett stickprov med fyra observationer från varje behandling så borde alla 16 observationerna vara väl blandade om nollhypotesen är sann medan de olika behandlingarna borde vara separerade om det finns skillnad någonstans.

Det är alltid bra att rita upp observationerna för att se så att det inte finns värden som är extrema på något sätt. Om man använder Lådagram i Excel för dessa observationer får man inte ett riktigt lådagram utan istället en bra bild av observationerna för de olika behandlingarna.



Figur 5.10.

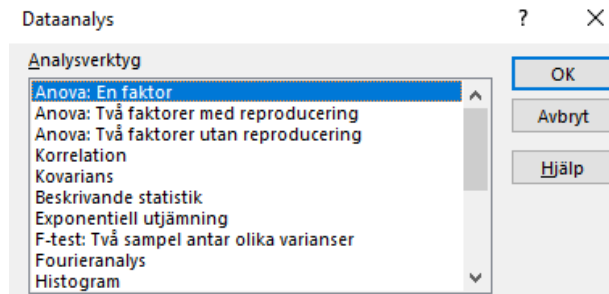
Lådagram ger *inte* ett lådagram men en bra bild av observationerna.

Det är omöjligt att med blotta ögat avgöra om det finns en så stor skillnad mellan behandlingarna att den är statistiskt signifikant (= tydlig). Det finns ju en tendens till att B har liten skörd och att C har en ganska stor skörd, men frågan är om det kan vara slumpen som har orsakat skillnaden.

För att testa om det är en skillnad kan man använda Dataanalys i Excel, men för att avgöra *var* skillnaderna är så går det inte med den funktionen. Excel vill ha de olika behandlingarna i olika kolumner för att kunna göra beräkningen, och det är tvärtemot de råd som finns i Avsnitt 3.3 och också tvärtemot hur andra program vill ha datamaterialet. I detta fall skall datamaterialet ligga så här i Excel:

	A	B	C	D
1	A	B	C	D
2	25	21	24	23
3	23	23	23	25
4	20	20	24	21
5	21	19	26	25

För att göra envägs variansanalys väljer du Dataanalys och "Anova; En faktor"



Resultatet blir en tabell med beskrivande statistik för de olika behandlingarna och en så kallad variansanalystabell (= ANOVA-tabell).

Anova: En faktor						
SAMMANFATTNING						
Grupper	Antal	Summa	Medelvärde	Varians		
A	4	89	22,25	4,916667		
B	4	83	20,75	2,916667		
C	4	97	24,25	1,583333		
D	4	94	23,5	3,666667		
ANOVA						
Variationsursprung	KvS	fg	MKv	F	p-värde	F-krit
Mellan grupper	28,1875	3	9,395833333	2,872611	0,080503	3,490295
Inom grupper	39,25	12	3,270833333			
Totalt	67,4375	15				

Här finns en hel del beteckningar och uttryck som inte är viktiga om du bara vill veta om det är en skillnad. De viktigaste sakerna i utskriften är:

- Medelvärde som du får automatiskt och kan använda om du vill rita ett stapeldiagram av resultatet.
- Varians som är Excels mått på spridningen. Använd funktionen =ROT på varianserna för att få fram standardavvikelserna som är det vanligaste valet på felstaplar i stapeldiagrammet av resultatet.
- F som är testvärde i den här analysen och detta värde bör vara 1 om nollhypotesen är sann. Här är $F = 2,872611$ och det är omöjligt att veta om detta är så mycket större än 1 så att du skall förkasta nollhypotesen.
- p-värde som är 0,081, och detta visar att F inte är tillräckligt stort för att ändra åsikt för den som tror att nollhypotesen är sann. Slutsatsen blir därför att nollhypotesen *inte* kan förkastas eftersom p-värdet är större än 5 %,

Det finns alltså inte någon signifikant skillnad i skörd mellan de olika behandlingarna. Skillnaden på 3,5 enheter mellan behandling C och behandling B är inte tillräckligt stor för att vi skall anse dem skilda åt. Slumpvariationen är så stor att de som har fått samma behandling kan ha en skillnad i tillväxt på upp till 5 enheter (för behandling A), och mot bakgrund av detta är det inte så stor skillnad mellan behandlingarnas medelvärden. ■

Om du nu hade fått en signifikant skillnad så hade det varit intressant att veta mellan vilka behandlingar det är en signifikant skillnad. Variansanalysen kan inte ge besked om detta, och den vanligaste metoden för att bestämma vilka behandlingar som skiljer sig från varandra finns beskriven i Avsnitt 5.13 om Tukeys test.

5.9 Variansanalys av ett blockförsök

Fördelen med att lägga ut försöket som ett blockförsök är att du minskar slumpvariationen när du för ihop liknande enheter i samma block och gör jämförelsen inom blocken. Det är bra att ha en liten slumpvariation och enkelt uttryckt beror det på att det då är lättare att se skillnader mellan behandlingarna. Som ett exempel på ett blockförsök finns här ett sortförsök med sex sorter, dvs. behandlingar, och fyra replikat, dvs. block.

Exempel 5.5.

För att undersöka om skördarna skiljer sig åt mellan sex olika sorter så görs ett blockförsök med fyra block. Inom varje block randomiserades sorterna med försöksplanen

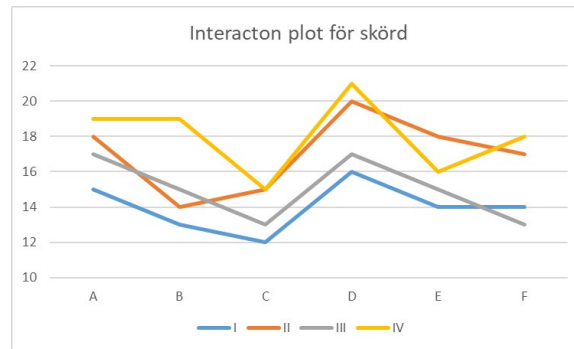
Block	Sort					
I	A	C	B	E	F	D
II	D	F	A	B	E	C
III	C	E	F	D	A	B
IV	B	A	D	C	F	E

När du sedan skall räkna på och redovisa rutskördarna i *kg* per ruta ordnar man sorterna i ordning. Siffrorna har i detta fall avrundats till heltal för att underlätta, men så får du självklart inte göra i ett riktigt försök.

Vid beräkningar för hand behöver du ha tillgång till medelvärdena för varje sort och medelvärdena för varje block. Det är därför som rutskördarna har ställts upp som i tabellen även om beräkningarna här görs med hjälp av dator.

Block	A	B	C	D	E	F	Medelvärde av block
I	15	13	12	16	14	14	14,00
II	18	14	15	20	18	17	17,00
III	17	15	13	17	15	13	15,00
IV	19	19	15	21	16	18	18,00
Medelvärde av sort	17,25	15,25	13,75	18,50	15,75	15,50	16,00

Genom att rita upp materialet som i Figur 5.11 med hjälp av ett linjediagram går det att se om det finns några klara sorteffekter och/eller blockeffekter. Genom att se om linjerna är parallella går det också att se om sorteffekterna är desamma över alla blocken. Testet utförs lite senare i det här avsnittet efter att den generella modellen har formulerats.



Figur 5.11.

Det finns en viss tendens till att sorterna skiljer sig åt och att denna skillnad är ganska samstämmig över de olika blocken.

Den generella modellen i ord är

$$\text{respons} = \text{blockeffekt} + \text{behandlingseffekt} + \text{slumpfel}$$

och idén är nu att dela upp den totala variationen i tre delar,

- en del som beror på att blocken är olika,
- en del som beror på att behandlingarna är olika,
- en del som beror på slumpen.

Därefter jämförs den del som beror på behandlingarna med den del som beror på slumpen för att se om behandlingseffekten dominerar över slumpvariationen.

Nollhypotesen är att responsen i medeltal är lika för alla behandlingarna, eller alternativt uttryckt att behandlingseffekten inte skiljer sig åt för de olika behandlingarna. Detta skulle i ord kunna uttryckas som

$$H_0: \text{"ingen skillnad mellan behandlingarnas medelvärden"}$$

Mothypotesen är att medelvärdet av responsen för de olika behandlingarna åtminstone i någon behandling skiljer sig från de övriga, eller alternativt att skillnaderna i respons mellan behandlingarna beror på mer än slumpmässiga orsaker skrivs

$$H_1: \text{"det är en skillnad mellan behandlingarnas medelvärden"}$$

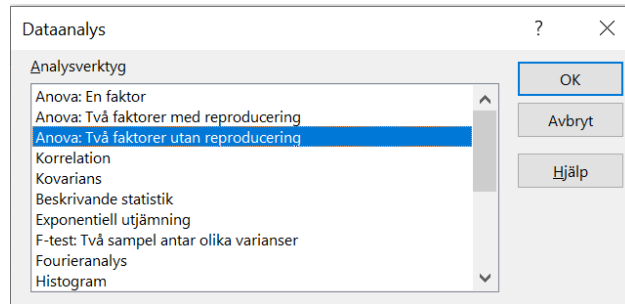
Det går att göra detta test i Excel men det fordrar att du lägger in data på det sätt som vi *inte* har rekommenderat tidigare. Med detta sätt att lägga in materialet är du begränsad om du vill flytta observationerna till ett annat program.

Exempel 5.5 (forts).

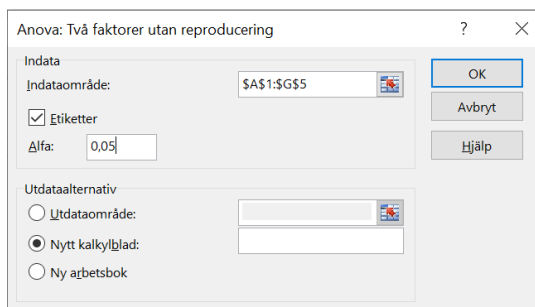
Lägg in observationerna så som de ligger i tabellen, dvs.

	A	B	C	D	E	F	G
1 Block	A	B	C	D	E	F	
2 I	15	13	12	16	14	14	
3 II	18	14	15	20	18	17	
4 III	17	15	13	17	15	13	
5 IV	19	19	15	21	16	18	

och använd Dataanalys och menyn



Markera tabellen, inklusive raden med sorter och kolumnen med block och kryssa i rutan Etiketter.



Resultatet blir både beskrivande statistik och en variansanalystabell.

Anova: Två faktorer utan reproducering						
<i>SAMMANFATTNING</i>	<i>Antal</i>	<i>Summa</i>	<i>Medelvärde</i>	<i>Varians</i>		
I	6	84	14	2		
II	6	102	17	4,8		
III	6	90	15	3,2		
IV	6	108	18	4,8		
A	4	69	17,25	2,91667		
B	4	61	15,25	6,91667		
C	4	55	13,75	2,25		
D	4	74	18,5	5,66667		
E	4	63	15,75	2,91667		
F	4	62	15,5	5,66667		
ANOVA						
<i>Variationsursprung</i>	<i>KvS</i>	<i>fg</i>	<i>MKv</i>	<i>F</i>	<i>p-värde</i>	<i>F-krit</i>
Rader	60	3	20	15,7895	6,6E-05	3,28738
Kolumner	55	5	11	8,68421	0,00049	2,90129
Fel	19	15	1,266666667			
Totalt	134	23				

Här finns medelvärden för behandlingar³⁰ med respektive varianser om du vill göra ett stapeldiagram.

Variansanalystabellen delar enligt tidigare upp variationen i tre delar.

Variation mellan block. Denna är ett mått på den variation (ojämnheter), som finns mellan blocken. I detta försök kan den knappast betecknas som obetydlig. Skörden

³⁰ Medelvärdena för block finns här också även om dessa inte är lika intressanta.

i Block I är i medeltal 14 kg per ruta medan den ligger på 17 kg per ruta i Block II. Om försöksplatsen varit absolut jämn skulle alla blockskördarna blivit desamma. I ett sådant fall skulle också p-värdet för block blivit större än 0,05.

Variation mellan behandlingarna. Denna är ett mått på variationen mellan behandlingarna, dvs. i detta fall skördeskillnaderna mellan sorternablocken. Ju större skillnad mellan sorterna, desto större blir värdet på F och ju mindre blir p-värdet. Kolumnen har p-värdet 0,00049. Detta p-värde visar att $F = 8,68$ är för långt ifrån 1 för att vi skall tro på nollhypotesen. Det går att förkasta nollhypotesen på signifikansnivån 0,1 % eller med *** signifikans.

Slumpvariationen. Kvar finns nu en rest som beror på den variation (ojämnhet) som döljer sig inom behandlingar eller inom block och som inte går att förklara med skillnader mellan behandlingar eller block.

Idén med F är liksom i det fullständigt randomiserade försöket tidigare att om nollhypotesen är sann (om det inte är någon skillnad mellan sorterna) så kommer F att hålla sig kring 1, och om mothypotesen är sann (om det är skillnad mellan sorterna) så kommer F troligen att vara större. I det här fallet är alltså F så stort och p-värdet så litet att det alldeles säkert är en skillnad mellan sorterna.

I datorutskriften finns det också ett testvärde F och ett p -värde för blocken. Dessa värden har ofta ett begränsat intresse eftersom syftet med experimentet var att undersöka om det var skillnad mellan behandlingarna. I vissa sammanhang används p-värdet för blocket för att se om det var motiverat att göra en blockindelning, och i detta fall var ju p-värdet så litet som 0,000066, vilket visar att det var en väldigt stor skillnad mellan blocken.

Men vilka behandlingar är det som skiljer sig från varandra? En rimlig tanke är att det finns en signifikant skillnad mellan behandling D med högst skörd och behandling C med lägst skörd, men det kanske finns fler signifikanta skillnader? Avsnitt 5.12 och Avsnitt 5.13 beskriver närmare hur du testar detta, men det går *inte* att göra med den här funktionen i Excel. ■

5.10 Variansanalys med flera faktorer

I våra exempel hittills har bara ingått en behandlingsfaktor. I Exempel 5.4 var det olika behandlingar, i Exempel 5.5 var behandlingarna olika sorter. Det är inte ovanligt att man vill studera två eller fler faktorer i samma försök och se hur de samspelar. Exempel 5.6 är ett exempel på detta med två faktorer, kväve och växtskydd.

Här skall vi väldigt kort beskriva hur man gör när det är mer än en faktor, och speciellt diskutera begreppet samspel (eng. interaction).

Samspel (= interaktion) innebär, att summan av faktorernas effekt när de verkar tillsammans, avviker från summan av effekterna när de verkar var för sig. Detta uttrycks som att det finns ett samspel. Exempel 5.6 har kväve på tre nivåer, 100, 140, 180 kg N/ha och växtskydd på två nivåer, obehandlat respektive behandlat enligt ett fastställt program. I försök med det här upplägget är det inte ovanligt att det finns samspel som yttrar sig i att skördeökningen med växtskydd är större vid

höga kvävegivor än vid låga eller inget *N*. Förklaringen ligger i att höga kvävegivor gynnar vissa skadegörare³¹.

Exempel 5.6

Försöksplan och skörd i *dt/ha* från ett tvåfaktoriellt försök med kväve och växtskydd i höstvetete. Blockförsök med fyra replikat.

Kväve	100 kg N	140 kg N	180 kg N	100 kg N	140 kg N	180 kg N
Växtskydd	Obehandlat	Obehandlat	Obehandlat	Behandlat	Behandlat	Behandlat
Block/Behandling	A	B	C	D	E	F
I	76,0	79,2	78,4	82,8	85,2	85,2
II	74,4	77,6	80,0	83,6	86,0	89,6
III	76,8	79,6	80,8	80,8	87,6	91,2
IV	77,2	78,0	78,8	83,2	89,2	90,0

Principen i analysen bygger liksom tidigare på att för de olika variationsorsakerna räkna fram testvärden som betecknas *F* och ur dessa *p*-värden.

För att se skillnaden mellan en försöksplan med en faktor och två faktorer så börjar vi med att analysera försöket utan att ta hänsyn till att det är två faktorer. Det är helt enkelt sex behandlingar A-F enligt tabellen ovan och variansanalystabellen går att beräkna med hjälp av Excel på samma sätt som vi gjorde i Avsnitt 5.8. Variansanalystabellen blir nu

ANOVA						
Variationsursprung	KvS	fg	Mkv	F	p-värde	F-krit
Rader	11,25333	3	3,751111111	1,634392	0,223553	3,287382
Kolumner	507,4933	5	101,4986667	44,22386	1,88E-08	2,901295
Fel	34,42667	15	2,295111111			
Totalt	553,1733	23				

Ur tabellen kan man utläsa att det är en klar skillnad mellan de sex behandlingarna på signifikansnivån 0,1 % eftersom *p*-värdet är 0,000000188, men det går inte att veta om det är kvävet eller växtskyddsbehandlingen som har orsakat skillnaden.

För att dela upp i kväve och växtskydd och se om det finns något samspel så använder vi här Minitab för att göra beräkningen. I den här analysen delar vi upp behandlingen i tre delar,

- en del som handlar om kvävet,
- en del som handlar om växtskydd,
- en del som handlar om samspelet mellan kväve och växtskydd.

Med uppdelningen i dessa delar så blir det en större variansanalystabell:

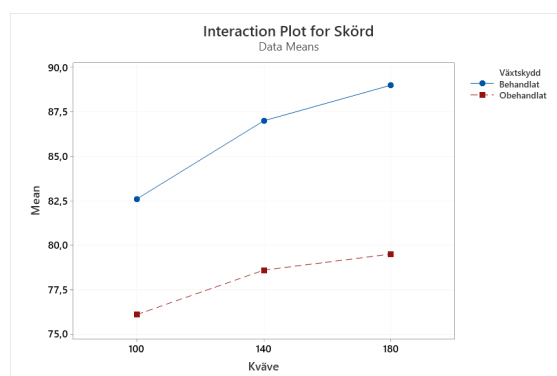
³¹ På svenska skulle det kanske också beskrivas som en synergieffekt.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Block	3	11,253	3,751	1,63	0,224
Kväve	2	101,373	50,687	22,08	0,000
Växtskydd	1	396,907	396,907	172,94	0,000
Kväve*Växtskydd	2	9,213	4,607	2,01	0,169
Error	15	34,427	2,295		
Total	23	553,173			

Lägg här speciellt märke till att de rader som betecknas Block, Error och Total är helt oförändrade och vad som har hänt är att raden med Behandling³² har delats upp i tre delar. Den rad som betecknas Kväve*Växtskydd är det som kallas samspel eller interaktion³³ (eng. interaction).

Kolumnen P-value visar att det inte finns något signifikant samspel mellan kväve och växtskydd (P är ju större än 0,05), det finns en signifikant effekt av kväve ($P \leq 0,001$) och en signifikant effekt av växtskydd ($P \leq 0,001$). Att samspelen inte är signifikant innebär att det *inte* finns bevis för att kvävet och växtskyddet samverkar, kvävet verkar fungera likadant oberoende av om växtskyddsbehandling utförs eller inte. Resultatet av detta illustreras kanske allra bäst av en "Interaction Plot".



Figur 5.12.

Illustration av samspelen mellan kväve och växtskydd.

Ur figuren går det att utläsa att den bästa skörden erhålls vid högsta dosen kväve och med växtskyddsbehandling. Att det inte finns något samspel innebär att de båda linjerna i stort sett är parallella och det som antydde i början har alltså *inte* kunnat påvisas, att växtskyddet skulle ha större effekt på högre kvävenivåer. Det signifikanta resultatet för växtskydd innebär att det är en skillnad i höjdlängd mellan linjerna och det signifikanta resultatet för kväve innebär att linjerna inte är horisontella, dvs. att det finns en skillnad mellan de olika kvävenivåerna.

Detta exempel har alltså visat styrkan med ett faktoriellt försök, hade det analyserats med sex led så hade det inte gått att plocka ut effekterna av kväve och växtskydd så som vi gjorde i den senare analysen. ■

³² Kolumn i den tidigare utskriften.

³³ Notera att programmet R betecknar samspel med Kväve:Växtskydd medan övriga program använder Kväve*Växtskydd.

Ju fler faktorer som ingår i försöket (experimentet), desto mer komplicerad blir analysen. Mönstret är dock detsamma, nämligen uppdelning av kvadratsummorna på olika komponenter (faktorer). I Exempel 5.6 ingår två faktorer, som ger en enkel samspelseffekt av typen Faktor 1*Faktor 2. Med tre faktorer erhålls även samspelseffekter av typen Faktor 1*Faktor 2*Faktor 3. Om nu signifikanta samspelseffekter erhålls är det viktigt att man kan finna en trovärdig biologisk förklaring till dem.

5.11 Eftertester i variansanalys

Variansanalysen anger bara om det är en signifikant skillnad eller inte och du får inte besked om mellan vilka av de enskilda leden som det är skillnad. För att få reda på detta gör man så kallade eftertester³⁴ (eng. post hoc test). Det finns många olika metoder, men två av de vanligaste är de som beskrivs här. Fishers metod räknar ut *LSD*-värdet och Tukeys metod räknar ut Tukey's Honestly Significant Difference, förkortat *HSD*. Då man använder dessa metoder underlättar det betydligt om man har balanserade försök, och därför är det så i de exempel som finns här.

Fishers metod kritiserar med all rätt för att den inte tar hänsyn till hur många parvisa jämförelser det är. Om nollhypotesen att det inte är någon skillnad är sann så växer risken att göra fel från 5 % till något större värde och signifikansnivån är inte längre 5 %. Med sex försöksled blir det väldigt stor chans att testet visar på skillnader även om det inte finns några. Det finns två sätt som brukar rekommenderas för att komma runt detta problem, men ingen av dem är perfekt:

- Använd bara Fishers metod på jämförelser som du har bestämt dig för innan försöket började och jämför inte alla par bara för att hitta p-värden under 5 %.
- Använd bara eftertester med Fishers metod om det är en signifikant skillnad i variansanalysen. Är inte p-värdet för behandling under 5 % i variansanalysen så avslutar du beräkningen och säger att det inte fanns några skillnader.

Det finns ett speciellt uttryck, "p-hacking", när du gör många tester som kanske inte alls var med i den ursprungliga hypotesen. Håller du på tillräckligt länge så kommer du förr eller senare att hitta ett signifikant resultat som kanske går att publicera.

Tukeys metod tar hänsyn till det här problemet och korrigerar för det. Tukeys metod tillåter därför test av alla parvisa jämförelser mellan behandlingar.

För att visa på problemet så återvänder vi till Exempel 5.4, där p-värdet på 0,081 visade att det *inte* fanns några signifikanta skillnader mellan de fyra behandlingarna.

³⁴ Dessa tester av skillnader kallas ofta "post-hoc tests" även på svenska, men här används beteckningen eftertester.

Exempel 5.4 (forts.)

Här redovisas bara utskrifterna, hur de närmare skall tolkas följer senare i avsnittet.

I Minitab finns både Fishers och Tukeys metod men först bekräftar vi att variansanalysstabellen ger p-värdet 0,081 också i Minitab:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	28,19	9,396	2,87	0,081
Error	12	39,25	3,271		
Total	15	67,44			

Tabellerna med bokstäver säger att Fishers metod hittar en skillnad mellan behandling C och behandling B eftersom de inte har samma bokstav i kolumnen Grouping, medan Tukeys metod har samma bokstav hela vägen och därmed *inte* hittar någon skillnad.

Grouping Information Using the Fisher LSD Method and 95% Confidence

Factor	N	Mean	Grouping
C	4	24,250	A
D	4	23,500	A B
A	4	22,25	A B
B	4	20,750	B

Means that do not share a letter are significantly different.

Grouping Information Using the Tukey Method and 95% Confidence

Factor	N	Mean	Grouping
C	4	24,250	A
D	4	23,500	A
A	4	22,25	A
B	4	20,750	A

Means that do not share a letter are significantly different.

Mot bakgrund av att p-värdet i variansanalysen var större än 5 % hade det nog varit tveksamt att publicera en rapport som sa att det var en skillnad mellan behandling C och behandling B trots att Fishers metod sa att det var så. ■

Detta var mer av karaktären "varnande exempel" och vi använder nu istället blockförsöket i Exempel 5.5 och förklarar lite mer detaljerat hur metoderna går till.

I Exempel 5.5 var det en väldigt tydlig skillnad mellan behandlingarna men vi vill veta var de signifikanta skillnaderna finns.

Ur variansanalysen går det att få en uppskattning av den gemensamma standardavvikelsen σ , och tillsammans med hur många observationer det är i försöket används detta för att avgöra vilka behandlingar som är signifikant skilda från varandra.

Det går att förvänta sig att Fishers metod ger fler signifikanta skillnader än Tukeys metod och därför är det lockande att använda den metoden i praktiken.

Trots alla varningar används därför ibland Fishers metod vid redovisning av fältförsök, i varje fall inom växtskydd och växtodling. Att metoden ger "för många" signifikanta resultat påpekas ganska sällan och följderna blir att resultaten omedvetet tolkas på ett felaktigt sätt.

Ett skäl till att inte använda Tukeys metod är att den kan ta till alltför mycket, och med många behandlingar så kan det leda till att du inte kan se några skillnader även om variansanalystabellen har ett p -värde som är under 5 %. Det finns därför en mängd alternativ "mellan" Fisher's och Tukey's metod men de behandlas inte här.

Ibland redovisas värdet på *LSD* och *HSD* under kolumnen med medelvärden från försöken utan att ange hur behandlingarna har grupperat sig. Läsaren av rapporten får sedan själv räkna ut vilka behandlingar som är signifikant skilda åt. Om du då inte vet hur metoden fungerar är det självklart en uppenbar risk att det blir fel-tolkningar och därför är det viktigt att veta grundidén med metoderna.

5.12 Fishers metod

Lägg märke till att många avråder från att använda Fishers metod men att den ändå är med här eftersom den används ganska ofta i praktiken, kanske speciellt när det är många behandlingar i försöket.

Fishers metod bygger på en utveckling av t -testet där man använder standardavvikelsen från variansanalystabellen för att uppskatta variationen i datamaterialet.

Genom att inte göra t -test som för två stickprov i Avsnitt 5.3 och Avsnitt 5.4 utan istället använda den här metoden som anknyter till variansanalysen så går det att göra testet lite mer effektivt.

Om stickproven är lika stora (alla har n observationer per led) så går det att få fram ett värde, kallat Minsta signifikanta differens eller *LSD* (eng. Least Significant Difference). Enligt denna metod är två behandlingar signifikant skilda åt om skillnaden i medelvärde överstiger värdet på *LSD*.

Det kan vara bra att veta att detta är grunden för hur du får den tabell med bokstäver som brukar publiceras i artiklarna. I de flesta fall behöver du inte bry dig om *LSD* eftersom datorn automatiskt ger tabellen med bokstäver.

Det spelar inte någon roll om det är en envägs variansanalys eller ett blockförsök, i båda fallen kan man få ett värde på *LSD* som anger hur stor skillnaden mellan medelvärdena skall vara för att anses som signifikant.

LSD kan relativt enkelt beräknas för hand med hjälp av en tabell men här antar vi att datorn kan leverera värdet så att vi slipper handräkna.

Exempel 5.5 (forts.)

Med $LSD = 1,70$ så är skillnader större än detta värde signifikant skilda från varandra. I datorprogrammen och i artiklar är det en tabell med bokstäver som illustrerar när det är en skillnad och det har ibland visat sig att många har svårt att tolka dessa tabeller.

Resultatet från Minitab blir tabellen

Grouping Information Using Fisher LSD Method and 95% Confidence

Sort	N	Mean	Grouping
D	4	18,50	A
A	4	17,25	A B
E	4	15,75	B C
F	4	15,50	C
B	4	15,25	C D
C	4	13,75	D

Means that do not share a letter are significantly different.

Texten under tabellen berättar vad det handlar om, de sorter som *inte* har en gemensam bokstav är signifikant skilda åt. I det här fallet betyder det t ex att sort D har högst medelvärde men är inte signifikant skild från sort A eftersom båda dessa har en gemensam bokstav, A, i kolumnen Grouping. Eftersom vi redan har fått reda på att $LSD = 1,70$ så stämmer det ju också med att $18,50 - 17,25 = 1,25 < 1,70$. Däremot är det en signifikant skillnad mellan behandling D och behandling E eftersom de *inte* har någon gemensam bokstav (bokstaven A finns inte i raden för behandling E i grouping som bara har B och C där). Det går också att motivera genom att $18,50 - 15,75 = 2,75 > 1,70$. Det går att motivera alla bokstäverna i kolumnen Grouping på detta sätt och helt göra tabellen för hand om man får reda på LSD, men vi stoppar här.

Det vanligaste sättet att redovisa resultatet är med en tabell som utskriften från Minitab, men med små bokstäver för grupptillhörigheten. Om det är ett balanserat försök skriver man ut n i tabelltexten och inte i tabellen.

Behandling	Medelvärde
D	18,50 a
A	17,25 ab
E	15,75 bc
F	15,50 c
B	15,25 cd
C	13,75 cd

Ett alternativ till tabellen är att göra en figur som illustreras i figur i nästa avsnitt.

Vid denna typ av tabeller står ofta angivet samma text som under tabellen men med signifikansnivån angiven, dvs. "medelvärden som inte har en gemensam bokstav är signifikant skilda åt på signifikansnivån 0,05". ■

5.13 Tukeys metod

Eftersom varje test av skillnader i Fishers metod är på signifikansnivån 5 % så kommer man förr eller senare att hitta "falska" signifikanta skillnader. Detta beror på att testets idé är att i 1 fall av 20 så kommer du att hitta en skillnad även om den inte finns. Letar du därför tillräckligt länge så kommer du att hitta skillnader, och detta är alltså det som kallas p-hacking.

För att undvika detta används istället alternativ till Fishers metod och det vanligaste alternativet är Tukeys metod. Detta test är baserat på *HSD* (Honestly Significant Difference³⁵), och är till idén väldigt likt Fishers *LSD*. Mycket enkelt uttryckt

³⁵ Ibland kallas även den siffra man får fram med Tukey's test för *LSD* istället för *HSD*.

handlar det om att undvika att anklagas för p-hacking genom att öka värdet på *LSD* så att det inte blir för många signifikanta skillnader. Det går därför att vara helt säker på att *HSD* alltid är större än eller lika stort som *LSD*, och att Fishers metod alltid ger minst lika många signifikanta parvisa skillnader som Tukeys metod.

Värdet på *HSD* kan beräknas för hand med hjälp av tabell men här antar vi att datorn levererar värdet så att vi slipper handräkna.

Exempel 5.5 (forts).

Eftersom det är sex sorter i försöket kan man göra femton parvisa jämförelser mellan leden. För att garantera att vi inte får för hög signifikansnivå och annonserar skillnader som egentligen inte finns så höjs värdet från $LSD = 1,70$ till $HSD = 2,59$. Det är alltså en ganska stor skillnad på *LSD* och *HSD* men det beror på att det är många sorter som jämförs. Antalet par av jämförelser växer snabbt med antalet sorter och med sortförsök med riktigt många sorter så kommer Tukeys metod att kompensera så mycket att det inte blir några signifikanta skillnader även om variansanalystabellen visar på en skillnad. Fishers metod används därför ibland i sortförsök trots alla dess nackdelar.

Med Tukeys metod behövs alltså en skillnad i medelvärden på mer än 2,59 för att två sorter skall anses som signifikanta skilda åt.

Utskriften av detta från Minitab blir

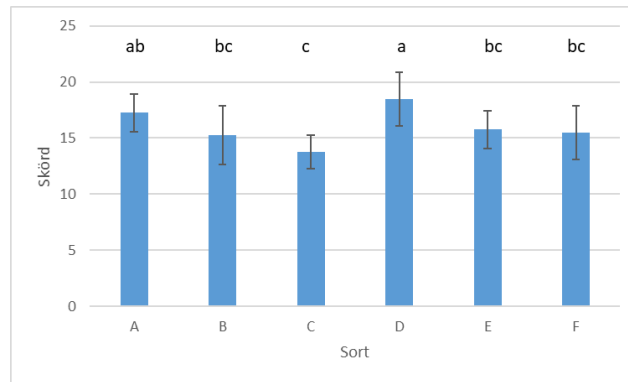
Grouping Information Using the Tukey Method and 95% Confidence

Behandling	N	Mean	Grouping
D	4	18,50	A
A	4	17,25	A B
E	4	15,75	B C
F	4	15,50	B C
B	4	15,25	B C
C	4	13,75	C

Means that do not share a letter are significantly different.

Det blir lite mindre skillnader jämfört med Fishers metod. Skillnaden mellan behandling E och behandling C är $15,75 - 13,75 = 2,00$ och det gör att den skiljer sig åt i Fishers metod men i Tukeys metod har dessa behandlingar den gemensamma bokstaven C och är därför inte skilda åt. Det stämmer också med att $2,00 < 2,59$.

I Avsnitt 5.12 om Fishers test visades hur du illustrerar med tabell, och här använder vi därför en figur som också är ett vanligt sätt att illustrera resultatet. Även om resultatet kom från Minitab så är det lättare att göra figuren i Excel så som illustrerades i Avsnitt 4.7.



Figur 5.13.

Staplarna illustrerar medelvärde och standardavvikelse och bokstäverna är resultatet av Tukeys metod³⁶.



Det är inte alltid helt lätt att göra en sammanfattande tolkning av dessa bokstavstabeller, speciellt om det är mer än en bokstav på raderna. Om däremot det bara finns en bokstav för varje behandling så blir det en lättare tolkning, eftersom metoden då har delat materialet i olika åtskilda grupper.

5.14 Ytterligare metoder

Detta kompendium innehåller de vanligaste metoderna men det finns självklart andra metoder som också ofta används. Några av dessa metoder finns i Minitabkompendiet, t.ex.

- icke-parametriska metoder som används om det inte är normalfördelat,
- chi-två-test som används när observationerna är frekvenser
- regression och korrelation när man vill jämföra två variabler.

³⁶ Bokstäverna ovanför staplarna är inlagda för hand som en textruta.