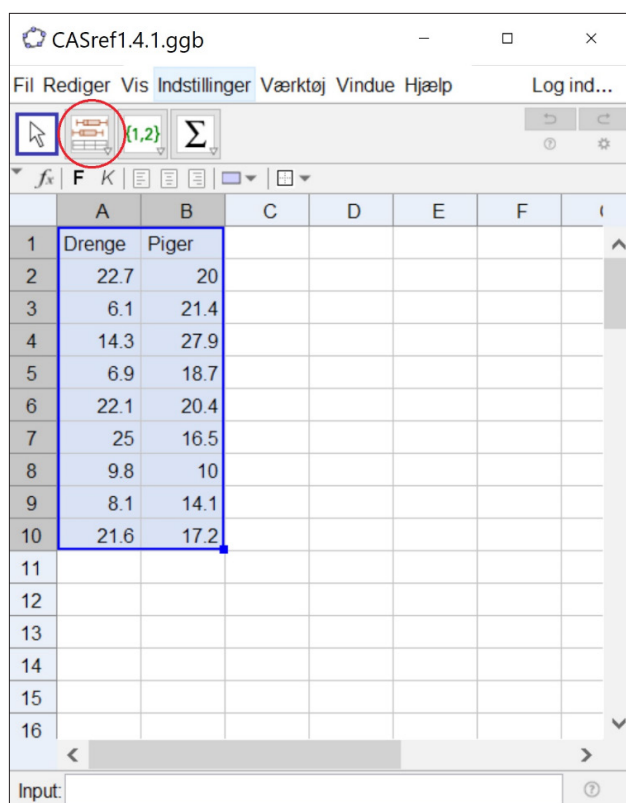


t-test i GeoGebra

I dette eksempel vises beregningen af en t -test i GeoGebra. Der bruges data fra 'Regn med biologi' om pigers og drenges fedtprocenter. Beregning af t -test minder meget om analysen foretaget i supplerende materiale (1.5), hvor CAS-værktøjet blev brugt til at lave bokplot og residualplot. Fremgangsmåden er derfor den samme som ved fremstilling af et bokplot. Der er bare tilføjet en matematisk beregning til sidst af middelværdi, spredning, spredning for en stikprøve, minimalværdi, 1. kvartil, median, 3. kvartil og maksimalværdi. Ud fra de matematiske beregninger kan også vælges en t -test.

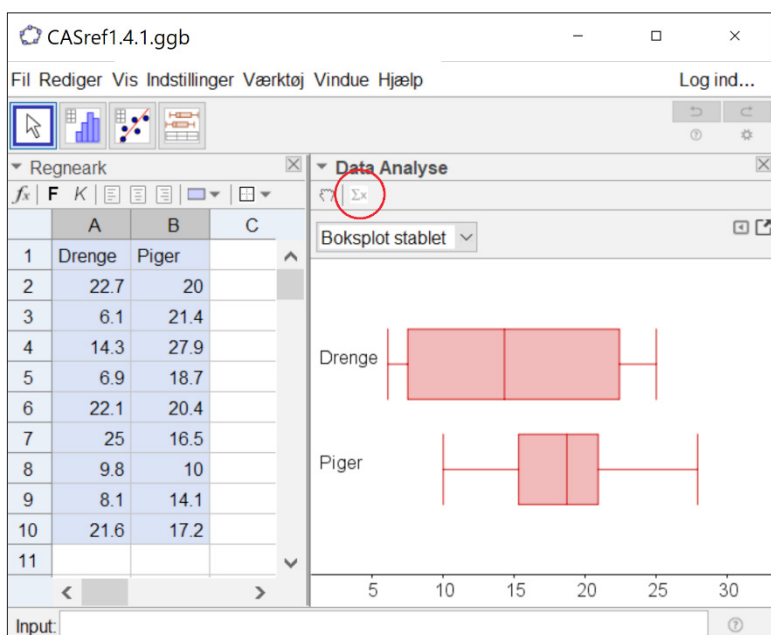
Eksempel med data fra 1.s

- 1) Start GeoGebra og vælg 'Regneark'.
- 2) Indtast data om fedtprocenter fra 1.s. Husk at bruge punktum som decimaladskiller. Resultatet ses i figur 1.
- 3) Marker alle data med musen.
- 4) Klik herefter på nederste højre hjørne i 'analyseværktøjer' hvor der er en trekant. Analyseværktøjer er markeret med rød ring i figur 1.
- 5) Vælg 'Flervariabelanalyse'. Klik på tandhjulet ude til højre over data og vælg 'Brug sidehoved som titel'. Klik på 'Analyser'. Der kommer et bokplot frem. Se figur 2.



	A	B	C	D	E	F	G
1	Dreng	Piger					
2	22.7	20					
3	6.1	21.4					
4	14.3	27.9					
5	6.9	18.7					
6	22.1	20.4					
7	25	16.5					
8	9.8	10					
9	8.1	14.1					
10	21.6	17.2					
11							
12							
13							
14							
15							
16							

Figur 1. Indtastning af data i GeoGebra.



Figur 2. Resultatet af flervariabelanalyse.

Herefter skal man bruge data fra flervariabelanalysen til at lave en *t*-test.

- 6) Klik på 'Σx' (markeret med rød cirkel på figur 2).
- 7) Under boksplottet kommer en deskriptiv statistik frem som vist i figur 23 i bogen.
- 8) I stedet for 'Statistik' kan på drop-down-menuen vælges 'T-test, forskel i middel'. Så testes om middelværdien er større i den ene gruppe end i den anden.
- 9) I dette tilfælde vil drengene stå øverst og pigerne nederst i sammenligningen fordi drengene er indtastet i venstre kolonne og pigerne i højre kolonne i regnearket. Det giver en negativ *t*-værdi, men man kan med drop-down-menu ændre drengene til at stå nederst og piger til at stå øverst. For altid at få positive *t*-værdier skal gruppen med den største middelværdi stå øverst. Resultatet af *t*-testen ses i figur 3.

T test, forskel i middel				
<input type="checkbox"/> Samenvægtet				
Nulhypotese: $\mu_1 - \mu_2 = 0$				
Alternativ hypotese: $\mu_1 - \mu_2 \neq 0$				
		Middel	s	n
Sample 1	Piger	18.4667	4.9995	9
Udvalg 2	Drengene	15.1778	7.6888	9
Forskel	P	t	SE	df
3.2889	0.3005	1.0758	3.0571	13.739

Figur 3. Resultat af *t*-test for beregning af fedtprocenter i 1.s.

Det ses af figur 3 at man tester om middelværdien for de to grupper μ_1 og μ_2 er ens, ved at måle om forskellen mellem dem er 0. Er det ikke korrekt, er den alternative hypotese at forskellen mellem de to gruppers middelværdi er forskellig fra 0. Resultatet viser at t -testen har en t -værdi på 1,0758 og en p -værdi på 0,3005. Dermed er der 0,3005 sandsynlighed for at resultatet er tilfældigt. Det ligger over den vedtagne værdi på 0,05, og man kan konstatere at 0-hypotesen ikke kan forkastes, og der dermed ikke er forskel på middelværdierne af de to gruppers fedtprocenter. Det skyldes at stikprøven er lille. Derfor kan man lave samme fremgangsmåde, men skrive data ind for både 1.s og 1.r med drenge og piger i hver sin kolonne og så sammenligne alle eleverne fordelt på køn.

Noter

Der er både nogle regnetekniske og nogle forståelsesmæssige problemstillinger man skal iagttage, således at man ikke laver fejlkonklusioner.

p -værdi

Værdien bruges normalt til at acceptere eller forkaste en nulhypotese ($H_{(0)}$) ud fra en på forhånd fastsat værdi. Man bruger typisk 0,05 eller 0,01 som værdi for at forkaste en nulhypotese om at der ikke er forskel på to datasæt. Man skal dog være opmærksom på at der ikke nødvendigvis er en forskel, selvom man får en p -værdi på 0,05. En p -værdi på 0,05 kan betyde to ting. Den ene er at der rent faktisk er forskel på de to datasæt, og den anden betydning er at man har været uheldig med data, og derfor ved et tilfælde har fået et resultat af den statistiske test der viser en forskel, som ikke har en faglig begrundelse. Ved en p -værdi på 0,05 vil det ske i 5 % af tilfældene. Det er betydningen af en p -værdi på 0,05. Og i statistik findes der ingen redskaber til at sige om det er det ene eller det andet forhold, der gør sig gældende. Derfor kan gentagelser af forsøg eller meget store datasæt være med til at få en lav p -værdi og et mere sandsynligt resultat.

t -test, parrede forskelle og t -test, forskel i middel

Man kan også i GeoGebra beregne t -test for parrede forskelle. Det er vigtigt at den kun bruges, når man har samme gruppe at sammenligne med før og efter et givet forsøg. Vil man fx måle om træning giver et højere kondital, kan man med fordel bruge ' t -test, parrede forskelle'. Sofies kondital sammenlignes så med hendes eget tal før og efter træningsprogrammet, Søren med Sørens eget tal, Alis tal før og efter og så videre. Derved kan det testes på et sandsynlighedsniveau, fx $p < 0,05$ at træningsprogrammet virker. Når data tages ind laves en kolonne der hedder 'Før træning' og en kolonne der hedder 'Efter træning'. Står Sofies data for 'Før træning' i A2 i regnearket skal hendes data for 'Efter træning' stå i samme række i B-kolonnen svarende til B2. Sørens tal kan fx stå i A3 og B3, og Alis tal kan stå i A4 og B4 osv. Så vil den parrede t -test fungere. Vil man derimod teste om drenge har højere kondital end piger, kan tallene bare skrives ind i tilfældig rækkefølge i de to grupper, og t -testen laves ved at bruge ' t -test forskel i middel'.